

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
9 August 2001 (09.08.2001)

PCT

(10) International Publication Number
WO 01/57085 A2

- (51) International Patent Classification⁷: C07K 14/705, 16/28, C12N 5/10, C12Q 1/68, A61K 38/17, A01K 67/027, G01N 33/53
- (74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (21) International Application Number: PCT/US01/03455
- (22) International Filing Date: 1 February 2001 (01.02.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/180,093 2 February 2000 (02.02.2000) US
60/182,045 11 February 2000 (11.02.2000) US
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GI, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant (*for all designated States except US*): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US).
- Published:**
— *without international search report and to be republished upon receipt of that report*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 01/57085 A2

(54) Title: G-PROTEIN COUPLED RECEPTORS

(57) Abstract: The invention provides human G-protein coupled receptors (GCREC) and polynucleotides which identify and encode GCREC. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of GCREC.

G-PROTEIN COUPLED RECEPTORS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of G-protein coupled receptors
5 and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative,
neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and
viral infections, and in the assessment of the effects of exogenous compounds on the expression of
nucleic acid and amino acid sequences of G-protein coupled receptors.

10 BACKGROUND OF THE INVENTION

Signal transduction is the general process by which cells respond to extracellular signals.
Signal transduction across the plasma membrane begins with the binding of a signal molecule, e.g., a
hormone, neurotransmitter, or growth factor, to a cell membrane receptor. The receptor, thus activated,
triggers an intracellular biochemical cascade that ends with the activation of an intracellular target
15 molecule, such as a transcription factor. This process of signal transduction regulates all types of cell
functions including cell proliferation, differentiation, and gene transcription. The G-protein coupled
receptors (GPCRs), encoded by one of the largest families of genes yet identified, play a central role in
the transduction of extracellular signals across the plasma membrane. GPCRs have a proven history of
being successful therapeutic targets.

20 GPCRs are integral membrane proteins characterized by the presence of seven hydrophobic
transmembrane domains which together form a bundle of antiparallel alpha (α) helices. GPCRs range
in size from under 400 to over 1000 amino acids (Strosberg, A.D. (1991) Eur. J. Biochem. 196:1-10;
Coughlin, S.R. (1994) Curr. Opin. Cell Biol. 6:191-197). The amino-terminus of a GPCR is
extracellular, is of variable length, and is often glycosylated. The carboxy-terminus is cytoplasmic and
25 generally phosphorylated. Extracellular loops alternate with intracellular loops and link the
transmembrane domains. Cysteine disulfide bridges linking the second and third extracellular loops
may interact with agonists and antagonists. The most conserved domains of GPCRs are the
transmembrane domains and the first two cytoplasmic loops. The transmembrane domains account, in
part, for structural and functional features of the receptor. In most cases, the bundle of α helices forms
30 a ligand-binding pocket. The extracellular N-terminal segment, or one or more of the three extracellular
loops, may also participate in ligand binding. Ligand binding activates the receptor by inducing a
conformational change in intracellular portions of the receptor. In turn, the large, third intracellular
loop of the activated receptor interacts with a heterotrimeric guanine nucleotide binding (G) protein
complex which mediates further intracellular signaling activities, including the activation of second

messengers such as cyclic AMP (cAMP), phospholipase C, and inositol triphosphate, and the interaction of the activated GPCR with ion channel proteins. (See, e.g., Watson, S. and S. Arkinstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 2-6; Bolander, F.F. (1994) Molecular Endocrinology, Academic Press, San Diego CA, pp. 162-176; Baldwin, J.M. (1994) Curr. Opin. Cell Biol. 6:180-190.)

5 GPCRs include receptors for sensory signal mediators (e.g., light and olfactory stimulatory molecules); adenosine, γ -aminobutyric acid (GABA), hepatocyte growth factor, melanocortins, neuropeptide Y, opioid peptides, opsins, somatostatin, tachykinins, vasoactive intestinal polypeptide family, and vasopressin; biogenic amines (e.g., dopamine, epinephrine and norepinephrine, histamine, 10 glutamate (metabotropic effect), acetylcholine (muscarinic effect), and serotonin); chemokines; lipid mediators of inflammation (e.g., prostaglandins and prostanoids, platelet activating factor, and leukotrienes); and peptide hormones (e.g., bombesin, bradykinin, calcitonin, C5a anaphylatoxin, endothelin, follicle-stimulating hormone (FSH), gonadotropic-releasing hormone (GnRH), neurokinin, and thyrotropin-releasing hormone (TRH), and oxytocin). GPCRs which act as receptors for stimuli 15 that have yet to be identified are known as orphan receptors.

The diversity of the GPCR family is further increased by alternative splicing. Many GPCR genes contain introns, and there are currently over 30 such receptors for which splice variants have been identified. The largest number of variations are at the protein C-terminus. N-terminal and cytoplasmic loop variants are also frequent, while variants in the extracellular loops or transmembrane domains are 20 less common. Some receptors have more than one site at which variance can occur. The splicing variants appear to be functionally distinct, based upon observed differences in distribution, signaling, coupling, regulation, and ligand binding profiles (Kilpatrick, G.J. et al. (1999) Trends Pharmacol. Sci. 20:294-301).

GPCRs can be divided into three major subfamilies: the rhodopsin-like, secretin-like, and 25 metabotropic glutamate receptor subfamilies. Members of these GPCR subfamilies share similar functions and the characteristic seven transmembrane structure, but have divergent amino acid sequences. The largest family consists of the rhodopsin-like GPCRs, which transmit diverse extracellular signals including hormones, neurotransmitters, and light. Rhodopsin is a photosensitive GPCR found in animal retinas. In vertebrates, rhodopsin molecules are embedded in membranous 30 stacks found in photoreceptor (rod) cells. Each rhodopsin molecule responds to a photon of light by triggering a decrease in cGMP levels which leads to the closure of plasma membrane sodium channels. In this manner, a visual signal is converted to a neural impulse. Other rhodopsin-like GPCRs are directly involved in responding to neurotransmitters. These GPCRs include the receptors for adrenaline (adrenergic receptors), acetylcholine (muscarinic receptors), adenosine, galanin, and glutamate (N-

methyl-D-aspartate/NMDA receptors). (Reviewed in Watson, S. and S. Arkinstall (1994) The G-Protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 7-9, 19-22, 32-35, 130-131, 214-216, 221-222; Habert-Ortoli, E. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9780-9783.)

The galanin receptors mediate the activity of the neuroendocrine peptide galanin, which
5 inhibits secretion of insulin, acetylcholine, serotonin and noradrenaline, and stimulates prolactin and growth hormone release. Galanin receptors are involved in feeding disorders, pain, depression, and Alzheimer's disease (Kask, K. et al. (1997) Life Sci. 60:1523-1533). Other nervous system rhodopsin-like GPCRs include a growing family of receptors for lysophosphatidic acid and other
10 lysophospholipids, which appear to have roles in development and neuropathology (Chun, J. et al. (1999) Cell Biochem. Biophys. 30:213-242).

The largest subfamily of GPCRs, the olfactory receptors, are also members of the rhodopsin-like GPCR family. These receptors function by transducing odorant signals. Numerous distinct olfactory receptors are required to distinguish different odors. Each olfactory sensory neuron expresses only one type of olfactory receptor, and distinct spatial zones of neurons expressing distinct receptors
15 are found in nasal passages. For example, the RA1c receptor which was isolated from a rat brain library, has been shown to be limited in expression to very distinct regions of the brain and a defined zone of the olfactory epithelium (Raming, K. et al. (1998) Receptors Channels 6:141-151). However, the expression of olfactory-like receptors is not confined to olfactory tissues. For example, three rat genes encoding olfactory-like receptors having typical GPCR characteristics showed expression
20 patterns not only in taste and olfactory tissue, but also in male reproductive tissue (Thomas, M.B. et al. (1996) Gene 178:1-5).

Members of the secretin-like GPCR subfamily have as their ligands peptide hormones such as secretin, calcitonin, glucagon, growth hormone-releasing hormone, parathyroid hormone, and vasoactive intestinal peptide. For example, the secretin receptor responds to secretin, a peptide
25 hormone that stimulates the secretion of enzymes and ions in the pancreas and small intestine (Watson, supra, pp. 278-283). Secretin receptors are about 450 amino acids in length and are found in the plasma membrane of gastrointestinal cells. Binding of secretin to its receptor stimulates the production of cAMP.

Examples of secretin-like GPCRs implicated in inflammation and the immune response
30 include the EGF module-containing, mucin-like hormone receptor (Emr1) and CD97 receptor proteins. These GPCRs are members of the recently characterized EGF-TM7 receptors subfamily. These seven transmembrane hormone receptors exist as heterodimers in vivo and contain between three and seven potential calcium-binding EGF-like motifs. CD97 is predominantly expressed in leukocytes and is markedly upregulated on activated B and T cells (McKnight, A.J. and S. Gordon

(1998) J. Leukoc. Biol. 63:271-280).

The third GPCR subfamily is the metabotropic glutamate receptor family. Glutamate is the major excitatory neurotransmitter in the central nervous system. The metabotropic glutamate receptors modulate the activity of intracellular effectors, and are involved in long-term potentiation (Watson, *supra*, p.130). The Ca^{2+} -sensing receptor, which senses changes in the extracellular concentration of calcium ions, has a large extracellular domain including clusters of acidic amino acids which may be involved in calcium binding. The metabotropic glutamate receptor family also includes pheromone receptors, the GABA_B receptors, and the taste receptors.

Other subfamilies of GPCRs include two groups of chemoreceptor genes found in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*, which are distantly related to the mammalian olfactory receptor genes. The yeast pheromone receptors STE2 and STE3, involved in the response to mating factors on the cell membrane, have their own seven-transmembrane signature, as do the cAMP receptors from the slime mold *Dictyostelium discoideum*, which are thought to regulate the aggregation of individual cells and control the expression of numerous developmentally-regulated genes.

GPCR mutations, which may cause loss of function or constitutive activation, have been associated with numerous human diseases (Coughlin, *supra*). For instance, retinitis pigmentosa may arise from mutations in the rhodopsin gene. Furthermore, somatic activating mutations in the thyrotropin receptor have been reported to cause hyperfunctioning thyroid adenomas, suggesting that certain GPCRs susceptible to constitutive activation may behave as protooncogenes (Parma, J. et al. (1993) Nature 365:649-651). GPCR receptors for the following ligands also contain mutations associated with human disease: luteinizing hormone (precocious puberty); vasopressin V_2 (X-linked nephrogenic diabetes); glucagon (diabetes and hypertension); calcium (hyperparathyroidism, hypocalcemia, hypercalcemia); parathyroid hormone (short limbed dwarfism); β_3 -adrenoceptor (obesity, non-insulin-dependent diabetes mellitus); growth hormone releasing hormone (dwarfism); and adrenocorticotropin (glucocorticoid deficiency) (Wilson, S. et al. (1998) Br. J. Pharmacol. 125:1387-1392; Stadel, J.M. et al. (1997) Trends Pharmacol. Sci. 18:430-437). GPCRs are also involved in depression, schizophrenia, sleeplessness, hypertension, anxiety, stress, renal failure, and several cardiovascular disorders (Horn, F. and G. Vriend (1998) J. Mol. Med. 76:464-468).

In addition, within the past 20 years several hundred new drugs have been recognized that are directed towards activating or inhibiting GPCRs. The therapeutic targets of these drugs span a wide range of diseases and disorders, including cardiovascular, gastrointestinal, and central nervous system disorders as well as cancer, osteoporosis and endometriosis (Wilson, *supra*; Stadel, *supra*). For example, the dopamine agonist L-dopa is used to treat Parkinson's disease, while a dopamine antagonist is used to treat schizophrenia and the early stages of Huntington's disease. Agonists and antagonists of

adrenoceptors have been used for the treatment of asthma, high blood pressure, other cardiovascular disorders, and anxiety; muscarinic agonists are used in the treatment of glaucoma and tachycardia; serotonin 5HT_{1D} antagonists are used against migraine; and histamine H₁ antagonists are used against allergic and anaphylactic reactions, hay fever, itching, and motion sickness (Horn, supra).

- 5 Recent research suggests potential future therapeutic uses for GPCRs in the treatment of metabolic disorders including diabetes, obesity, and osteoporosis. For example, mutant V2 vasopressin receptors causing nephrogenic diabetes could be functionally rescued in vitro by co-expression of a C-terminal V2 receptor peptide spanning the region containing the mutations. This result suggests a possible novel strategy for disease treatment (Schöneberg, T. et al. (1996) EMBO J. 15:1283-1291).
- 10 Mutations in melanocortin-4 receptor (MC4R) are implicated in human weight regulation and obesity. As with the vasopressin V2 receptor mutants, these MC4R mutants are defective in trafficking to the plasma membrane (Ho, G. and R.G. MacKenzie (1999) J. Biol. Chem. 274:35816-35822), and thus might be treated with a similar strategy. The type 1 receptor for parathyroid hormone (PTH) is a GPCR that mediates the PTH-dependent regulation of calcium homeostasis in the bloodstream. Study
- 15 of PTH/receptor interactions may enable the development of novel PTH receptor ligands for the treatment of osteoporosis (Mannstadt, M. et al. (1999) Am. J. Physiol. 277:F665-F675).

- The chemokine receptor group of GPCRs have potential therapeutic utility in inflammation and infectious disease. (For review, see Locati, M. and P.M. Murphy (1999) Annu. Rev. Med. 50:425-440.) Chemokines are small polypeptides that act as intracellular signals in the regulation of leukocyte
- 20 trafficking, hematopoiesis, and angiogenesis. Targeted disruption of various chemokine receptors in mice indicates that these receptors play roles in pathologic inflammation and in autoimmune disorders such as multiple sclerosis. Chemokine receptors are also exploited by infectious agents, including herpesviruses and the human immunodeficiency virus (HIV-1) to facilitate infection. A truncated version of chemokine receptor CCR5, which acts as a coreceptor for infection of T-cells by HIV-1,
- 25 results in resistance to AIDS, suggesting that CCR5 antagonists could be useful in preventing the development of AIDS.

- The discovery of new G-protein coupled receptors and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, neurological, cardiovascular, gastrointestinal,
- 30 autoimmune/inflammatory, and metabolic disorders, and viral infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of G-protein coupled receptors.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, G-protein coupled receptors, referred to collectively as "GCREC" and individually as "GCREC-1," "GCREC-2," "GCREC-3," "GCREC-4," "GCREC-5," "GCREC-6," "GCREC-7," "GCREC-8," "GCREC-9," "GCREC-10," "GCREC-11," "GCREC-12," "GCREC-13," "GCREC-14," "GCREC-15," "GCREC-16," "GCREC-17," "GCREC-18," "GCREC-19," "GCREC-20," and "GCREC-21." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-21. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:22-42.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90%

sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) culturing a cell under conditions suitable for
5 expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid
10 sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

15 The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e)
20 an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of
25 SEQ ID NO:22-42, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide
30 in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as

an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional GCRC, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the

activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:22-42, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted

motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"GCREC" refers to the amino acid sequences of substantially purified GCREC obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of GCREC. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

- 5 An "allelic variant" is an alternative form of the gene encoding GCREC. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides.
- 10 Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

- "Altered" nucleic acid sequences encoding GCREC include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as GCREC or a polypeptide with at least one functional characteristic of GCREC. Included within this definition are
- 15 polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding GCREC, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding GCREC. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent GCREC.
- 20 Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of GCREC is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity
- 25 values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

- The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic
- 30 molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known

in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of GCREC. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by
5 directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind GCREC polypeptides can be prepared using intact polypeptides or using
10 fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize
15 the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the
20 protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as
25 phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring
30 nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic"

refers to the capability of the natural, recombinant, or synthetic GCREC, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

“Complementary” describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A “composition comprising a given polynucleotide sequence” and a “composition comprising a given amino acid sequence” refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution.

Compositions comprising polynucleotide sequences encoding GCREC or fragments of GCREC may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

“Consensus sequence” refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

“Conservative amino acid substitutions” are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
30	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
35	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val

	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
5	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

10

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

15 A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

20 A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

25 A "fragment" is a unique portion of GCRC or the polynucleotide encoding GCRC which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 30 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported 35 by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:22-42 comprises a region of unique polynucleotide sequence that

specifically identifies SEQ ID NO:22-42, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:22-42 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:22-42 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:22-42 and the region of SEQ ID NO:22-42 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-21 is encoded by a fragment of SEQ ID NO:22-42. A fragment of SEQ ID NO:1-21 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-21. For example, a fragment of SEQ ID NO:1-21 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-21. The precise length of a fragment of SEQ ID NO:1-21 and the region of SEQ ID NO:1-21 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search

Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at

<http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other

- 5 polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to
- 10 compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

- 15 *Open Gap: 5 and Extension Gap: 2 penalties*

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

- 20 Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences
- 25 shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences

- 30 that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions,

explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment
5 program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

10 Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

15 *Gap x drop-off: 50*

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for
20 example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to
25 describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid
30 sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity.

Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

- Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

- High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

- The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of GCREC which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of GCREC which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of GCREC. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of GCREC.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an GCREC may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by

cell type depending on the enzymatic milieu of GCREC.

"Probe" refers to nucleic acid sequences encoding GCREC, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical
5 labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

10 Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the
15 specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR
20 Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such
25 purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South
30 West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for

microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that
5 hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary
10 polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the
15 artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.
20 Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs).
25 Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and
30 other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing GCREC, nucleic acids encoding GCREC, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

- 5 The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope
- 10 A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

- The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which
- 15 they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

- "Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers,
- 20 microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

- "Transformation" describes a process by which exogenous DNA is introduced into a recipient
- 25 cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells"
- 30 includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic

acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in
5 vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention
10 into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999)
15 set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides
20 due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given
25 species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the
30 polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human G-protein coupled receptors (GCREC), the polynucleotides encoding GCREC, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, neurological, cardiovascular, gastrointestinal,
 5 autoimmune/inflammatory, and metabolic disorders, and viral infections.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte
 10 polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the
 15 polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where
 20 applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential
 25 phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable
 databases to which the analytical methods were applied.

30 As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention.

Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:22-42 or that distinguish between SEQ ID NO:22-42 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 5080262H1 is the identification number of an Incyte cDNA sequence, and LNODNOT11 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., SBSA02572V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g4589483_CD) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g5902227_030.edit is the identification number of a Genscan-predicted coding sequence, with g5902227 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses GCREC variants. A preferred GCREC variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the GCREC amino acid sequence, and which contains at least one functional or

structural characteristic of GCREC.

The invention also encompasses polynucleotides which encode GCREC. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42, which encodes GCREC. The polynucleotide sequences of
5 SEQ ID NO:22-42, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding GCREC. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least
10 about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding GCREC. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:22-42.
15 Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of GCREC.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding GCREC, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the
20 invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring GCREC, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode GCREC and its variants are generally capable of
25 hybridizing to the nucleotide sequence of the naturally occurring GCREC under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding GCREC or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which
30 particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding GCREC and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode GCREC and

GCREC derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding GCREC or any fragment thereof.

5 Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:22-42 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in
10 "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or
15 combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing
20 system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

25 The nucleic acid sequences encoding GCREC may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site-PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)
30 Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al.

(1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060).

- 5 Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a
10 GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

- When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library
15 does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

- Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-
20 specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments
25 which may be present in limited amounts in a particular sample.

- In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode GCREC may be cloned in recombinant DNA molecules that direct expression of GCREC, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent
30 amino acid sequence may be produced and used to express GCREC.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter GCREC-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic

oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of GCREC, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding GCREC may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, GCREC itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of GCREC, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active GCREC, the nucleotide sequences encoding GCREC or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding GCREC. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding GCREC. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding GCREC and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding GCREC and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding GCREC. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and

Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. 5 (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding GCRC. For example, routine cloning, 10 subcloning, and propagation of polynucleotide sequences encoding GCRC can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding GCRC into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro 15 transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of GCRC are needed, e.g. for the production of antibodies, vectors which direct high level expression of GCRC may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

20 Yeast expression systems may be used for production of GCRC. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; 25 Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of GCRC. Transcription of sequences encoding GCRC may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock 30 promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill,

New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding GCREC may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader
5 sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses GCREC in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

10 Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of
15 GCREC in cell lines is preferred. For example, sequences encoding GCREC can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a
20 selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase
25 genes, for use in *tk*⁻ and *aprt*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to
30 chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used.

These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding GCREC is inserted within a marker gene sequence, transformed cells containing sequences encoding GCREC can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding GCREC under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding GCREC and that express GCREC may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of GCREC using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on GCREC is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) *Serological Methods, a Laboratory Manual*, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) *Current Protocols in Immunology*, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) *Immunochemical Protocols*, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding GCREC include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding GCREC, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease

of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding GCREC may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein
5 produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode GCREC may be designed to contain signal sequences which direct secretion of GCREC through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the
10 inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities
15 (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding GCREC may be ligated to a heterologous sequence resulting in translation of a
20 fusion protein in any of the aforementioned host systems. For example, a chimeric GCREC protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of GCREC activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose
25 binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion
30 proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the GCREC encoding sequence and the heterologous protein sequence, so that GCREC may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled GCREC may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

GCREC of the present invention or fragments thereof may be used to screen for compounds that specifically bind to GCREC. At least one and up to a plurality of test compounds may be screened for specific binding to GCREC. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of GCREC, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which GCREC binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express GCREC, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing GCREC or cell membrane fractions which contain GCREC are then contacted with a test compound and binding, stimulation, or inhibition of activity of either GCREC or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with GCREC, either in solution or affixed to a solid support, and detecting the binding of GCREC to the compound.

Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

GCREC of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of GCREC. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for GCREC activity, wherein GCREC is combined with at least one test compound, and the activity of GCREC in the presence of a test compound is compared with the activity of GCREC in the absence of the test compound. A change in the activity of GCREC in the presence of the test compound is indicative of a compound that modulates the activity of GCREC. Alternatively, a test compound is

combined with an in vitro or cell-free system comprising GCREC under conditions suitable for GCREC activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of GCREC may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

- 5 In another embodiment, polynucleotides encoding GCREC or their mammalian homologs may be “knocked out” in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo
- 10 and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330).
- 15 Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.
- 20 Polynucleotides encoding GCREC may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).
- 25 Polynucleotides encoding GCREC can also be used to create “knockin” humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding GCREC is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with
- 30 potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress GCREC, e.g., by secreting GCREC in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists

between regions of GCREC and G-protein coupled receptors. Therefore, GCREC appears to play a role in cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections. In the treatment of disorders associated with increased GCREC expression or activity, it is desirable to decrease the expression or activity of GCREC. In the

5 treatment of disorders associated with decreased GCREC expression or activity, it is desirable to increase the expression or activity of GCREC.

Therefore, in one embodiment, GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC. Examples of such disorders include, but are not limited to, a cell proliferative

10 disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia,

15 gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy,

20 retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis,

25 tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia

30 gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease,

35 aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular

tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, 5 rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric 10 carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, 15 colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, 20 hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune 25 hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel 30 syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, 35 parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes,

obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and tongavirus.

5 In another embodiment, a vector capable of expressing GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified GCREC in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of GCREC may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those listed above.

15 In a further embodiment, an antagonist of GCREC may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GCREC. Examples of such disorders include, but are not limited to, those cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections, described above. In one aspect, an antibody which specifically binds GCREC may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express GCREC.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding GCREC may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GCREC including, but not limited to, those described above.

25 In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of GCREC may be produced using methods which are generally known in the art. In particular, purified GCREC may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind GCREC. Antibodies to GCREC may

also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

- 5 For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with GCREC or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, 10 polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

- It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to GCREC have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or 15 fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of GCREC amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

- Monoclonal antibodies to GCREC may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited 20 to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

- In addition, techniques developed for the production of "chimeric antibodies," such as the 25 splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce GCREC-specific single 30 chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in

the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for GCREC may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between GCREC and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering GCREC epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for GCREC. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of GCREC-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple GCREC epitopes, represents the average affinity, or avidity, of the antibodies for GCREC. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular GCREC epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the GCREC-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of GCREC, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of GCREC-antibody

complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding GCREC, or any fragment
 5 or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding GCREC. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding GCREC.
 10 (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g.,
 15 Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other
 20 systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding GCREC may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency
 25 (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene*
 30 *Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g.,

against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in GCREC expression or regulation causes disease, the expression of GCREC from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in GCREC are treated by constructing mammalian expression vectors encoding GCREC and introducing these vectors by mechanical means into GCREC-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of GCREC include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). GCREC may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding GCREC from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these

standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to GCREC expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding GCREC under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding GCREC to cells which have one or more genetic abnormalities with respect to the expression of GCREC. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding GCREC to target cells which have one or more genetic abnormalities with respect to the expression of GCREC. The use of herpes simplex virus (HSV)-based vectors may be

especially valuable for introducing GCREC to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 5 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this 10 patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of 15 herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding GCREC to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on 20 the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for GCREC into the alphavirus 25 genome in place of the capsid-coding region results in the production of a large number of GCREC-coding RNAs and the synthesis of high levels of GCREC in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application. 30 (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of GCREC into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases,

5 transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

10 Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding GCRC.

15 Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of
20 candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis.

25 Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding GCRC. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

30 RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and

wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding GCRC.

5 Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or
10 promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased GCRC expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding GCRC may be therapeutically useful, and in the treatment of disorders associated with decreased GCRC expression or activity, a compound which specifically promotes expression of the polynucleotide encoding GCRC may be therapeutically useful.

15 At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound
20 based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding GCRC is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding GCRC are assayed
25 by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding GCRC. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide
30 exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys.
35 Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a

combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

- 5 Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

- 15 An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of GCREC, antibodies to GCREC, and mimetics, agonists, antagonists, or inhibitors of GCREC.

- 20 The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

- Compositions for pulmonary administration may be prepared in liquid or dry powder form.
- 25 These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising GCREC or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, GCREC or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example GCREC or fragments thereof, antibodies of GCREC, and agonists, antagonists or inhibitors of GCREC, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art.

Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

5 In another embodiment, antibodies which specifically bind GCREC may be used for the diagnosis of disorders characterized by expression of GCREC, or in assays to monitor patients being treated with GCREC or agonists, antagonists, or inhibitors of GCREC. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for GCREC include methods which utilize the antibody and a label to detect GCREC
10 in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

 A variety of protocols for measuring GCREC, including ELISAs, RIAs, and FACS, are known
15 in the art and provide a basis for diagnosing altered or abnormal levels of GCREC expression. Normal or standard values for GCREC expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to GCREC under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of GCREC expressed in
20 subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

 In another embodiment of the invention, the polynucleotides encoding GCREC may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and
25 quantify gene expression in biopsied tissues in which expression of GCREC may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of GCREC, and to monitor regulation of GCREC levels during therapeutic intervention.

 In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding GCREC or closely related molecules may be used to
30 identify nucleic acid sequences which encode GCREC. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding GCREC, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the GCREC encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:22-42 or from genomic sequences including promoters, enhancers, and introns of the GCREC gene.

- 5 Means for producing specific hybridization probes for DNAs encoding GCREC include the cloning of polynucleotide sequences encoding GCREC or GCREC derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety
10 of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

- Polynucleotide sequences encoding GCREC may be used for the diagnosis of disorders associated with expression of GCREC. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis,
15 hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid,
20 penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating
25 diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis,
30 encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including

- mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis,
- 5 hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve,
- 10 mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal
- 15 carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis,
- 20 Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein
- 25 obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia,
- 30 asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polycystic disease-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's
- 35 thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis,

myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal
5 circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and tongavirus. The polynucleotide sequences encoding
10 GCREC may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered GCREC expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding GCREC may be useful in assays that
15 detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding GCREC may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control
20 sample then the presence of altered levels of nucleotide sequences encoding GCREC in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of
25 GCREC, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding GCREC, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified
30 polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the

patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding GCREC may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding GCREC, or a fragment of a polynucleotide complementary to the polynucleotide encoding GCREC, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of GCREC include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be
5 accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray
10 can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the
15 activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, GCREC, fragments of GCREC, or antibodies specific for GCREC
20 may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by
25 quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of
transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the
30 hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the

case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and

analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are
5 visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein
10 spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

15 A proteomic profile may also be generated using antibodies specific for GCREC to quantify the levels of GCREC expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoz, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of
20 methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J.
25 Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological
30 sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the

individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding GCREC may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial PI constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding GCREC on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is
5 valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may
10 also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, GCREC, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a
15 solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between GCREC and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are
20 synthesized on a solid substrate. The test compounds are reacted with GCREC, or fragments thereof, and washed. Bound GCREC is then detected by methods well known in the art. Purified GCREC can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

25 In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding GCREC specifically compete with a test compound for binding GCREC. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with GCREC.

In additional embodiments, the nucleotide sequences which encode GCREC may be used in any
30 molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are,

therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/180,093 and U.S. Ser. No. 60/182,045, are expressly incorporated by
5 reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database
10 (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and
15 ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was
20 isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the
25 recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column
30 chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells

including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision
5 using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of
10 distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using
15 PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation
20 such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).
25 Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997,
30 supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA

sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families.

- 5 See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length.
- 10 Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently
- 15 analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are
- 20 generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

- Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold
- 25 parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity
 - 30 between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:22-42. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative G-protein coupled receptors were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode G-protein coupled receptors, the encoded polypeptides were analyzed by querying against PFAM models for G-protein coupled receptors. Potential G-protein coupled receptors were also identified by homology to Incyte cDNA sequences that had been annotated as G-protein coupled receptors. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to

be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were
5 given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpr public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

10 **"Stretched" Sequences**

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis
15 to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the
20 public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of GCREC Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:22-42 were compared with
25 sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:22-42 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for
30 Genome Research (WIGR), and Génethon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, or human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-

arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

- Alternatively, polynucleotide sequences encoding GCREC are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one
- 5 of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories.
- 10 Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding GCREC. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ
- 15 GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of GCREC Encoding Polynucleotides

- Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was
- 20 synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.
- 25 Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

- High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$,
- 30 and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2:

94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:22-42 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is

specifically described, essentially the same procedure is used with larger nucleotide fragments.

Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ - 32 P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser

desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

5 **Tissue or Cell Sample Preparation**

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μ l 5X SSC/0.2% SDS.

20 **Microarray Preparation**

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110° C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average

concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).

Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water.

- 5 Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

- Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and
10 Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for
15 about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

- Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines
20 at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

- 25 In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is
30 typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

- The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that
35 location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples

from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

- 5 The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and
10 measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

- A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used
15 for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

- Sequences complementary to the GCREC-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring GCREC. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with
20 smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of GCREC. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the GCREC-encoding transcript.

25 XII. Expression of GCREC

- Expression and purification of GCREC is achieved using bacterial or virus-based expression systems. For expression of GCREC in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (lac)* hybrid
30 promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express GCREC upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of GCREC in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus

(AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding GCREC by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, GCREC is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from GCREC at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16). Purified GCREC obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

XIII. Functional Assays

GCREC function is assessed by expressing the sequences encoding GCREC at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of

fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in
5 expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of GCREC on gene expression can be assessed using highly purified populations
10 of cells transfected with sequences encoding GCREC and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression
15 of mRNA encoding GCREC and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of GCREC Specific Antibodies

GCREC substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to
20 immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the GCREC amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well
25 described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the
30 oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for anti-peptide and anti-GCREC activity by, for example, binding the peptide or GCREC to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring GCREC Using Specific Antibodies

Naturally occurring or recombinant GCREC is substantially purified by immunoaffinity

chromatography using antibodies specific for GCREC. An immunoaffinity column is constructed by covalently coupling anti-GCREC antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

- 5 Media containing GCREC are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of GCREC (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/GCREC binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and GCREC is collected.

10 **XVI. Identification of Molecules Which Interact with GCREC**

- Molecules which interact with GCREC may include agonists and antagonists, as well as molecules involved in signal transduction, such as G proteins. GCREC, or a fragment thereof, is labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) A fragment of GCREC includes, for example, a fragment comprising one or more of the
15 three extracellular loops, the extracellular N-terminal region, or the third intracellular loop. Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled GCREC, washed, and any wells with labeled GCREC complex are assayed. Data obtained using different concentrations of GCREC are used to calculate values for the number, affinity, and association of GCREC with the candidate ligand molecules.

- 20 Alternatively, molecules interacting with GCREC are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech). GCREC may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between
25 the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

- Potential GCREC agonists or antagonists may be tested for activation or inhibition of GCREC receptor activity using the assays described in sections XVII and XVIII. Candidate molecules may be selected from known GPCR agonists or antagonists, peptide libraries, or combinatorial chemical
30 libraries.

Methods for detecting interactions of GCREC with intracellular signal transduction molecules such as G proteins are based on the premise that internal segments or cytoplasmic domains from an orphan G protein-coupled seven transmembrane receptor may be exchanged with the analogous domains of a known G protein-coupled seven transmembrane receptor and used to identify the G-

proteins and downstream signaling pathways activated by the orphan receptor domains (Kobilka, B.K. et al. (1988) Science 240:1310-1316). In an analogous fashion, domains of the orphan receptor may be cloned as a portion of a fusion protein and used in binding assays to demonstrate interactions with specific G proteins. Studies have shown that the third intracellular loop of G protein-coupled
 5 seven transmembrane receptors is important for G protein interaction and signal transduction (Conklin, B.R. et al. (1993) Cell 73:631-641). For example, the DNA fragment corresponding to the third intracellular loop of GCREC may be amplified by the polymerase chain reaction (PCR) and subcloned into a fusion vector such as pGEX (Pharmacia Biotech). The construct is transformed into an appropriate bacterial host, induced, and the fusion protein is purified from the cell lysate by
 10 glutathione-Sepharose 4B (Pharmacia Biotech) affinity chromatography.

For in vitro binding assays, cell extracts containing G proteins are prepared by extraction with 50 mM Tris, pH 7.8, 1 mM EGTA, 5 mM MgCl₂, 20 mM CHAPS, 20% glycerol, 10 µg of both aprotinin and leupeptin, and 20 µl of 50 mM phenylmethylsulfonyl fluoride. The lysate is incubated on ice for 45 min with constant stirring, centrifuged at 23,000 g for 15 min at 4°C, and the
 15 supernatant is collected. 750 µg of cell extract is incubated with glutathione S-transferase (GST) fusion protein beads for 2 h at 4°C. The GST beads are washed five times with phosphate-buffered saline. Bound G subunits are detected by [³²P]ADP-ribosylation with pertussis or cholera toxins. The reactions are terminated by the addition of SDS sample buffer (4.6% (w/v) SDS, 10% (v/v) β-mercaptoethanol, 20% (w/v) glycerol, 95.2 mM Tris-HCl, pH 6.8, 0.01% (w/v) bromophenol blue).
 20 The [³²P]ADP-labeled proteins are separated on 10% SDS-PAGE gels, and autoradiographed. The separated proteins in these gels are transferred to nitrocellulose paper, blocked with blotto (5% nonfat dried milk, 50 mM Tris-HCl (pH 8.0), 2 mM CaCl₂, 80 mM NaCl, 0.02% NaN₃, and 0.2% Nonidet P-40) for 1 hour at room temperature, followed by incubation for 1.5 hours with Gα subtype selective antibodies (1:500; Calbiochem-Novabiochem). After three washes, blots are incubated with
 25 horseradish peroxidase (HRP)-conjugated goat anti-rabbit immunoglobulin (1:2000, Cappel, Westchester PA) and visualized by the chemiluminescence-based ECL method (Amersham Corp.).

XVII. Demonstration of GCREC Activity

An assay for GCREC activity measures the expression of GCREC on the cell surface. cDNA encoding GCREC is transfected into an appropriate mammalian cell line. Cell surface proteins are
 30 labeled with biotin as described (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using GCREC-specific antibodies, and immunoprecipitated samples are analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of GCREC expressed on the cell surface.

35 In the alternative, an assay for GCREC activity is based on a prototypical assay for

ligand/receptor-mediated modulation of cell proliferation. This assay measures the rate of DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding GCREC is added to quiescent 3T3 cultured cells using transfection methods well known in the art. The transiently transfected cells are then incubated in the presence of [³H]thymidine, a radioactive DNA precursor molecule. Varying amounts of GCREC ligand are then added to the cultured cells. Incorporation of [³H]thymidine into acid-precipitable DNA is measured over an appropriate time interval using a radioisotope counter, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold GCREC ligand concentration range is indicative of receptor activity. One unit of activity per milliliter is defined as the concentration of GCREC producing a 50% response level, where 100% represents maximal incorporation of [³H]thymidine into acid-precipitable DNA (McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York NY, p. 73.)

In a further alternative, the assay for GCREC activity is based upon the ability of GPCR family proteins to modulate G protein-activated second messenger signal transduction pathways (e.g., cAMP; Gaudin, P. et al. (1998) *J. Biol. Chem.* 273:4990-4996). A plasmid encoding full length GCREC is transfected into a mammalian cell line (e.g., Chinese hamster ovary (CHO) or human embryonic kidney (HEK-293) cell lines) using methods well-known in the art. Transfected cells are grown in 12-well trays in culture medium for 48 hours, then the culture medium is discarded, and the attached cells are gently washed with PBS. The cells are then incubated in culture medium with or without ligand for 30 minutes, then the medium is removed and cells lysed by treatment with 1 M perchloric acid. The cAMP levels in the lysate are measured by radioimmunoassay using methods well-known in the art. Changes in the levels of cAMP in the lysate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

To measure changes in inositol phosphate levels, the cells are grown in 24-well plates containing 1×10^5 cells/well and incubated with inositol-free media and [³H]myoinositol, 2 μ Ci/well, for 48 hr. The culture medium is removed, and the cells washed with buffer containing 10 mM LiCl followed by addition of ligand. The reaction is stopped by addition of perchloric acid. Inositol phosphates are extracted and separated on Dowex AG1-X8 (Bio-Rad) anion exchange resin, and the total labeled inositol phosphates counted by liquid scintillation. Changes in the levels of labeled inositol phosphate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

XVIII. Identification of GCREC Ligands

GCREC is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK

(Human Embryonic Kidney) 293 which have a good history of GPCR expression and which contain a wide range of G-proteins allowing for functional coupling of the expressed GCREC to downstream effectors. The transformed cells are assayed for activation of the expressed receptors in the presence of candidate ligands. Activity is measured by changes in intracellular second messengers, such as cyclic AMP or Ca^{2+} . These may be measured directly using standard methods well known in the art, or by the use of reporter gene assays in which a luminescent protein (e.g. firefly luciferase or green fluorescent protein) is under the transcriptional control of a promoter responsive to the stimulation of protein kinase C by the activated receptor (Milligan, G. et al. (1996) Trends Pharmacol. Sci. 17:235-237). Assay technologies are available for both of these second messenger systems to allow high throughput readout in multi-well plate format, such as the adenylyl cyclase activation FlashPlate Assay (NEN Life Sciences Products), or fluorescent Ca^{2+} indicators such as Fluo-4 AM (Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In cases where the physiologically relevant second messenger pathway is not known, GCREC may be coexpressed with the G-proteins $\text{G}_{\alpha 15/16}$ which have been demonstrated to couple to a wide range of G-proteins (Offermanns, S. and M.I. Simon (1995) J. Biol. Chem. 270:15175-15180), in order to funnel the signal transduction of the GCREC through a pathway involving phospholipase C and Ca^{2+} mobilization. Alternatively, GCREC may be expressed in engineered yeast systems which lack endogenous GPCRs, thus providing the advantage of a null background for GCREC activation screening. These yeast systems substitute a human GPCR and G_α protein for the corresponding components of the endogenous yeast pheromone receptor pathway. Downstream signaling pathways are also modified so that the normal yeast response to the signal is converted to positive growth on selective media or to reporter gene expression (Broach, J.R. and J. Thorner (1996) Nature 384 (supp.):14-16). The receptors are screened against putative ligands including known GPCR ligands and other naturally occurring bioactive molecules. Biological extracts from tissues, biological fluids and cell supernatants are also screened.

25

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention.

Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments.

30 Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7472033	1	7472033CD1	22	7472033CB1
7472034	2	7472034CD1	23	7472034CB1
7472035	3	7472035CD1	24	7472035CB1
7472036	4	7472036CD1	25	7472036CB1
7472037	5	7472037CD1	26	7472037CB1
7472039	6	7472039CD1	27	7472039CB1
7472040	7	7472040CD1	28	7472040CB1
4250893	8	4250893CD1	29	4250893CB1
6726656	9	6726656CD1	30	6726656CB1
7472062	10	7472062CD1	31	7472062CB1
7472067	11	7472067CD1	32	7472067CB1
7472072	12	7472072CD1	33	7472072CB1
7472074	13	7472074CD1	34	7472074CB1
7472077	14	7472077CD1	35	7472077CB1
7472082	15	7472082CD1	36	7472082CB1
7472128	16	7472128CD1	37	7472128CB1
7472134	17	7472134CD1	38	7472134CB1
7472136	18	7472136CD1	39	7472136CB1
7472142	19	7472142CD1	40	7472142CB1
7472171	20	7472171CD1	41	7472171CB1
7472172	21	7472172CD1	42	7472172CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	7472033CD1	g4062997	4.3e-38	Neuropeptide Y receptor type 4 [Sus scrofa]
2	7472034CD1	g6532001	3.6e-89	Odorant receptor S19 [Mus musculus]
3	7472035CD1	g4680260	2.5e-72	Odorant receptor S18 [Mus musculus]
4	7472036CD1	g4680260	2.9e-87	Odorant receptor S18 [Mus musculus]
5	7472037CD1	g32093	2.5e-81	Olfactory receptor HGNP07J [Homo sapiens]
6	7472039CD1	g3983374	9.3e-75	Olfactory receptor C6 [Mus musculus]
7	7472040CD1	g7707801	1.0e-177	G protein-coupled receptor C5L2 [Homo sapiens]
8	4250893CD1	g3341996	5.0e-65	Angiotensin/vasopressin receptor AII/AVP [Homo sapiens] (Mao, M. et al. (1998) Proc. Natl. Acad. Sci. USA 95(14):8175-8180)
9	6726656CD1	g3983408	7.4e-82	Olfactory receptor H7 [Mus musculus] (Krautwurst, D. et al. (1998) Cell 95(7):917-926)
10	7472062CD1	g3746443	7.6e-80	Olfactory receptor OR93Ch [Pan troglodytes] (Rouquier, S. et al. (1998) Hum. Mol. Genet. 7(9):1337-1345)
11	7472067CD1	g2358267	6.3e-23	Galanin receptor [Mus musculus] (Jacoby, A.S. et al. (1997) Genomics 45(3):496-508)
12	7472072CD1	g1336043	3.4e-68	HsOLF3 [Homo sapiens]
13	7472074CD1	g10732802	1.0e-47	Vomerolateral receptor 1 [Homo sapiens]
14	7472077CD1	g8118040	1.0e-168	Orphan G-protein coupled receptor [Homo sapiens]
15	7472082CD1	g2317704	2.2e-114	Olfactory receptor [Rattus norvegicus] (McClintock, T.S. et al. (1997) Brain Res. Mol. Brain Res. 2:59-68)
16	7472128CD1	g1246534	3.5e-82	Olfactory receptor 4 [Gallus gallus] (Leibovici, M. et al. (1996) Dev. Biol. 175(1):118-131)
17	7472134CD1	g4680254	1.4e-78	Odorant receptor S1 [Mus musculus] (Malnic, B. et al. (1999) Cell 96(5):713-723)
18	7472136CD1	g206810	1.3e-43	G-protein coupled receptor [Rattus norvegicus] (Ross, P.C. et al. (1990) Proc. Natl. Acad. Sci. USA 87:3052-3056)
19	7472142CD1	g5869918	1.0e-93	Olfactory receptor [Mus musculus] (Strotmann, J. et al. (1999) Gene 236(2):281-291)
20	7472171CD1	g5869925	1.6e-134	Olfactory receptor [Mus musculus] (Strotmann, J. et al. (1999) Gene 236(2):281-291)
21	7472172CD1	g3983374	3.9e-76	Olfactory receptor C6 [Mus musculus] (Krautwurst, D. et al. (1998) Cell 95(7):917-926)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7472033CD1	470	S131 T153 S24 S73 Y82	N2 N46 N51 N56	G-protein coupled receptor motif: V184-I200 Transmembrane domains: P80-Y105, L145-Y164, V214-Y234, Y271-Y291 7 transmembrane receptor (rhodopsin family) signature: L140-Y379 G-protein coupled receptor signatures: Y164-P203, Y280-Y291, R316-F342, S371-R387 G-protein coupled receptor signature: A175-A224 Neuropeptide Y receptor signatures: T153-G165, Q178-A193, R314-I326, M366-N375, L377-I390 Rhodopsin-like GPCR superfamily signatures: M84-S108, Q178-I200, G212-V233, F269-A290, S272-Y295, T321-Y345, Y361-R387 Probable G protein-coupled receptor signatures: I174-A191, G380-L393 T22D1.12 protein (similar to G-protein coupled receptors): PD061410: Y271-A415 G-protein coupled receptor: PD000009: L140-Y234	MOTIFS HMNER HMNER-PFAM BLIMPS-BLOCKS PROFILESKAN BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODROM BLAST-PRODROM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1					G-protein coupled receptor DM00013 P49683 54-350: L140-L394 DM00013 P50391 35-337: S141-L394 DM00013 P49146 44-341: M84-K388 DM00013 P25929 34-335: S141-Q391	BLAST-DOMO
2	7472034CD1	326	T276 T323 S123 T323	N57	G-protein coupled receptor motif: M125-I141 Transmembrane domains: W61-D67, M161-C184, T208-G230 7 transmembrane receptor (rhodopsin family) signature: G56-Y307 G-protein coupled receptor signatures: H105-P144, E247-S273, F266-H277, P299-R315 Olfactory receptor signatures: M74-K95, A192-E206, F253-V268 Rhodopsin-like GPCR superfamily signatures: W41-F62, W41-W65, M74-K95, I119-I141, G214-L237, A252-T276, V289-R315 Olfactory receptor PD000921: L181-I260 Putative G-protein coupled receptor RA1c PD170483: I260-F322 G-protein coupled receptor: DM00013 P23269 15-304: I32-V321 DM00013 P30955 18-305: L49-V321 DM00013 S29708 18-306: Q39-V321 DM00013 P23274 18-306: E37-V321	MOTIFS HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-PRODOM BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	7472035CD1	315	T49 T103 S172 S48 T144 S224		G-protein coupled receptor motif: V105-I121	MOTIFS
					Signal peptide: M1-A28	HMMER
					Transmembrane domain: Y30-S48	HMMER
					7 transmembrane receptor (rhodopsin family) signature: G36-Y288	HMMER-PFAM
					G-protein coupled receptor signatures: S85-P124, T227-S253, P280-R296	BLIMPS-BLOCKS
					G-protein coupled receptor signature: F97-R146	PROFILES CAN
					Olfactory receptor signatures: M54-K75, S172-D186, L233-L248	BLIMPS-PRINTS
					Angiotensin II receptor signature: I209-V220	BLIMPS-PRINTS
					EDG1 orphan receptor signature: A43-F57	BLIMPS-PRINTS
					Olfactory receptor PD000921: L161-V240	BLAST-PRODOM
4	7472036CD1	314	S110	N5 N44	G-protein coupled receptor: DM00013 P23274 18-306: I12-I299 DM00013 G45774 18-309: P13-E297 DM00013 S29708 18-306: H19-I299 DM00013 P23269 15-304: Q16-I299	BLAST-DOMO
					G-protein coupled receptor motif: M112-I128	MOTIFS
					Transmembrane domains: I27-L42, I199-L218	HMMER
					7 transmembrane receptor (rhodopsin family) signature: G43-Y294	HMMER-PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4					G-protein coupled receptor signatures: R92-P131, E234-S260, P286-R302	BLIMPS-BLOCKS
					G-protein coupled receptor signature: F104-A152	PROFILES CAN
					Olfactory receptor signatures: M61-K82, C179-D193, F240-T255, L278-V289	BLIMPS-PRINTS
					Melanocortin receptor signature: A53-L65	BLIMPS-PRINTS
5	7472037CD1	321	S75 S196 S201 S240 T299	N12 N50	Rhodopsin-like GPCR superfamily signatures: W28-Q52, M61-K82, L106-I128, I143-L164, G201-L224, A203-V227, I276-R302	BLIMPS-PRINTS
					Olfactory receptor PD000921: Y170-I247	BLAST-PRODOM
					Putative G-protein coupled receptor RALC PD170483: I247-F309	BLAST-PRODOM
					G-protein coupled receptor: DM00013 P23275 17-306: H24-I305 DM00013 G45774 18-309: P20-I305 DM00013 S29708 18-306: E23-V308 DM00013 P23269 15-304: E23-V308	BLAST-DMO
5	7472037CD1	321	S75 S196 S201 S240 T299	N12 N50	Transmembrane domains: P33-G49, V148-I171, L205-I229	HMMER
					7 transmembrane receptor (rhodopsin family) signature: G49-Y298	HMMER-PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5					G-protein coupled receptor signatures: Q98-P137, F208-F219, R243-R269, T290-R306 G-protein coupled receptor signature: F110-S159 Olfactory receptor signatures: M67-N88, F185-E199, F246-G261, V282-L293, T299-L313 Melanocortin receptor signature: V59-L71, I134-R145 Olfactory receptor PD000921: F176-L253 Olfactory receptor PD149621: T254-R315 G-protein coupled receptor: DM00013 P30954 29-316: S26-L309 DM00013 P23270 18-311: L31-L313 DM00013 P30955 18-305: P29-K311 DM00013 S29709 11-299: P29-L313 G-protein coupled receptor motif: S112-I128 Signal peptide: M1-A53 Transmembrane domains: M61-L84, A98-M120, S207-L229, F241-V263 7 transmembrane receptor (rhodopsin family) signature: G43-Y293 G-protein coupled receptor signatures: H92-P131, Q238-R264, T285-K301	BLIMPS-BLOCKS PROFILESKAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-PRODOM BLAST-DOMO
6	7472039CD1	331	S69 S24 T195 S235 S326	N6 N40 N67		MOTIFS SPSCAN HMMER HMMER-PFAM BLIMPS-BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6					G-protein coupled receptor signature: Y104-A149	PROFILESCAN
					Olfactory receptor signatures: M61-K82, F180-N194, F241-G256, V277-L288, A294-F308	BLIMPS-PRINTS
					Melanocortin receptor signature: A53-L65, I128-S139	BLIMPS-PRINTS
					Olfactory receptor PD000921: L168-I249	BLAST-PRODOM
					Olfactory receptor PD149621: V250-K310	BLAST-PRODOM
					G-protein coupled receptor: DM00013 P23267 20-309: P20-K305 DM00013 P23270 18-311: R26-K305 DM00013 P23274 18-306: F30-L304 DM00013 P30955 18-305: F30-K305	BLAST-DOMO
7	7472040CD1	337	S17 S323 S326 S194 T327 S333	N3	7 transmembrane receptor (rhodopsin family) signature: G52-D131, V153-F291	HMMER-PFAM
					G-protein coupled receptor signatures: W100-P139, F209-H220, W226-L252, H283-L299	BLIMPS-BLOCKS
					C5A-anaphylatoxin receptor signatures: V59-W72, C83-P93, L110-S121 P111-L123	BLIMPS-PRINTS
					Rhodopsin-like GPCR superfamily signatures: A37-G61, G69-L90, I114-A136, G150-I171, T201-L224, C231-V255, E273-L299	BLIMPS-PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7					Prostanoid EP1 receptor signature: W58-L73 G-protein coupled receptor: PD000009: W72-Y172 G-protein coupled receptor: DM00013 P30992 32-317: L39-L292 DM00013 P21730 31-315: V36-L303 DM00013 P30993 27-313: P33-P304 DM00013 A46525 31-317: P33-P304	BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO
8	4250893CD1	1473	S990 S37 S49 S99 S162 S258 S299 S368 S379 S400 S485 S509 S587 T747 T850 T866 T868 S923 T962 T993 S1047 S1064 T1075 S1313 S1371 T1394 S132 S144 S227 T457 S484 T543 T670 T831 T841 S893 S985 S1009 S1017 S1107 S1258 S1464 Y1311 Y1445	N727 N864 N1005	Transmembrane domain: G1216-L1237 Leucine Rich Repeats: N809-R833; L838-R862; T866-R890; K895-P922; S923-R947 RECEPTOR ANGIOTENSIN/VASOPRESSIN ATII/AVP VASOPRESSIN PD156095: M534-V704 ATP/GTP binding site (P-loop): G334-S341	HMMER HMMER-PFAM BLAST-PRODOM MOTIFS
9	6726656CD1	328	S17 T309	N23	Transmembrane domains: V43-I63; Y211-M231 7 transmembrane receptors (rhodopsin family) domain: G59-Y308 G-protein coupled receptors signature: S108-P147; S300-K316	HMMER HMMER-PFAM BLIMPS-BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9					G-protein coupled receptors signature: Y120-V166 Olfactory receptor signature: M77-K98; F195-S209; F256-T271; A292-L303; T309-F323 RECEPTOR OLFATORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L184-L263 G-PROTEIN COUPLED RECEPTORS DM00013 P23266 17-306: S36-L322 G-protein coupled receptors motif: S128-I144	PROFILESSCAN BLIMPS-PRINTS BLAST-PRODROM BLAST-DOMO MOTIFS
10	7472062CD1	384	T298 S34 S36 T69 T256 S10 T146 T205 T231 S258 T338 T359 Y378	N49 N73 N133 N223	Transmembrane domains: F96-L116; V268-I291; G310-I328 7 transmembrane receptor (rhodopsin family) domain: G109-Y358 G-protein coupled receptors signature: K158-P197; I350-K366 G-protein coupled receptors signature: F170-G215 Olfactory receptor signature: M127-K148; F245-E259; L306-I321; A342-L353; T359-V373	HMNER HMNER-PFAM BLIMPS-BLOCKS PROFILESSCAN BLIMPS-PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10					OLFACTORY RECEPTOR PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD149621: I316-V373 G-PROTEIN COUPLED RECEPTORS DM00013 S51356 18-307: L85-A368 G-protein coupled receptors motif: V178-V194 Transmembrane domain: A41-L61	BLAST- PRODOM BLAST-DOMO MOTIFS
11	7472067CD1	419	T234 S81 S102 T331 S358 S398 T177 T326 S370 T375	N10 N15 N79 N151	7 transmembrane receptor (rhodopsin family) domain: G53-A133; W155-F306 G-protein coupled receptors signature: W104-P143; I243-A269; I298-R314 Rhodopsin-like GPCR superfamily signature: I38-L62; I73-I94; T153-F174; V248-W272; I288-R314 G-PROTEIN COUPLED RECEPTORS DM00013 P47211 27-319: V49-K322 Transmembrane domains: L30-F47; I101-M118; M197-I214 7 transmembrane receptor (rhodopsin family) domain: G41-Y290 G-protein coupled receptors signature: G90-P129; L207-Y218; T282-W298	HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DOMO HMMER HMMER-PFAM BLIMPS-BLOCKS
12	7472072CD1	314	S67 S188 S227 S291	N5 N155		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12					G-protein coupled receptors signature: F102-S147 Rhodopsin-like GPCR superfamily signature: L26-R50; M59-K80; L104-V126; L199-L222; A237-V261; N272-W298 Olfactory receptor signature: M59-K80; F177-D191; V238-G253; V274-L285; S291-L305 RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: Y168-I245 G-PROTEIN COUPLED RECEPTORS DM00013 P23275 17-306; L26-L305 G-protein coupled receptors motif: A110-V126	PROFILERSCAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS
13	7472074CD1	254	T159 S185 S110 S111 T150 T178 S222	N154 N157	Signal peptide: M1-G61 Transmembrane domain: V127-F145; M182-S199; I235-V254 5-hydroxytryptamine 2A receptor signature: H48-A68 G-protein coupled receptors signature: Y73-R112; S226-F252 PHEROMONE RECEPTOR VN1 VN2 VN3 VN7 VN5 VN4 VN6 PD009900: R38-Y246	SPSCAN MOTIFS BLIMPS-PRINTS BLIMPS-BLOCKS BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	7472077CD1	362	S8 S268		<p>Transmembrane domains: S28-M48; F66-N85; V91-A109; C164-M181; L205-L224; W251-N278</p> <p>7 transmembrane receptor (metabotropic glutamate family) domain: W22-Q271 (Score: -137.1; E-value: 0.35)</p> <p>B1 bradykinin receptor signature: T37-Q56</p> <p>PROTEIN BRIDE OF SEVENLESS PRECURSOR TRANSMEMBRANE GLYCOPROTEIN VISION SIGNAL PD151485: V91-P260</p>	<p>HMME</p> <p>HMME-PFAM</p> <p>BLIMPS-PRINTS</p> <p>BLAST-PRODOR</p>
15	7472082CD1	370	S334 S42 S68 S127 S153 S326 T138 T147 S351	N65	<p>Transmembrane domains: F88-L108; W203-L228; L260-A283</p> <p>7 transmembrane receptor (rhodopsin family) domain: G101-Y350</p> <p>G-protein coupled receptors signature: N150-P189; V342-K358</p> <p>Olfactory receptor signature: M119-K140; Y237-S251; F298-G313; S334-L345; S351-L365</p> <p>RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L226-L306</p> <p>G-PROTEIN COUPLED RECEPTORS DM00013 S51356 18-307: L77-V361</p>	<p>HMME</p> <p>HMME-PFAM</p> <p>BLIMPS-BLOCKS</p> <p>BLIMPS-PRINTS</p> <p>BLAST-PRODOR</p> <p>BLAST-DOMO</p>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	7472128CD1	319	T2 S74 T85 S95 T201 S298	N274	Transmembrane domain: V204-C226 7 transmembrane receptor (rhodopsin family) domain: G48-Y297 G-protein coupled receptors signature: K97-P136; T214-Y225; I289-K305 G-protein coupled receptors signature: F110-A157 Olfactory receptor signature: M66-K87; Y184-N198 L245-G260; V281-L292; S298-L312 RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L173-L253 G-PROTEIN COUPLED RECEPTORS DM00013 S51356 18-307: P28-R310 G-protein coupled receptors motif: A117-I133 Signal peptide: M1-G41 Transmembrane domains: L29-V48; I207-L226; I238-L262 7 transmembrane receptor (rhodopsin family) domain: G41-Y290 G-protein coupled receptors signature: K90-P129; T282-K298	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILES SCAN BLIMPS-PRINTS BLAST-PRODOR BLAST-DOMO MOTIFS SPSCAN HMMER HMMER-PFAM BLIMPS-BLOCKS
17	7472134CD1	312	S67 S87 S204 S291	N5		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17					G-protein coupled receptors signature: Y102-I146 Rhodopsin-like GPCR superfamily signature: Y26-C50; M59-K80; F104-I126; D199-I222; Q272-K298 Olfactory receptor signature: M59-K80; V177-D191; I238-G253; L274-L285; S291-L305 RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L166-H244 G-PROTEIN COUPLED RECEPTORS DM00013 P23267 20-309: F17-K306 G-protein coupled receptors motif: T110-I126	PROFILES SCAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS
18	7472136CD1	321	S209 T94 T121 S281 S284	N2 N6 N16 N92	Transmembrane domain: Y109-F129; V229-Y245 7 transmembrane receptor (rhodopsin family) domain: G44-S78; L104-Y276 G-protein coupled receptors signature: V91-P130; R214-Y240; S268-S284 G-protein coupled receptors signature: L104-L149	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILES SCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18					Rhodopsin-like GPCR superfamily signature: V29-G53; F61-T82; M105-V127; L219-V243; F258-S284 GPROTEIN COUPLED RECEPTOR TRANSMEMBRANE GLYCOPROTEIN MAS PROTOONCOGENE G PROTEINCOUPLED PROBABLE PD013244: C175-E305 G-PROTEIN COUPLED RECEPTORS DM00013 p23749 38-307: V29-L287	BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO
19	7472142CD1	316	T49 S67 T228 S88 S290	N5 N42 N65	Transmembrane domains: F28-I48; Y102-D121; L199-I227 7 transmembrane receptor (rhodopsin family) domain: G41-Y289 G-protein coupled receptors signature: K90-P129; T281-K297 G-protein coupled receptors signature: Y102-V147 Visual pigments (opsins) retinal binding site: S262-H315 Rhodopsin-like GPCR superfamily signature: V26-I50; M59-T80; S104-I126; M198-L221; A236-K260; K271-K297 Olfactory receptor signature: M59-T80; F176-S190; F237-G252; I273-L284; S290-L304	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN PROFILESCAN BLIMPS-PRINTS BLIMPS-PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19					RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: I166-L244 G-PROTEIN COUPLED RECEPTORS DM00013 P23275 17-306: S18-G305 G-protein coupled receptors motif: T110-I126	BLAST- PRODOR MOTIFS
20	7472171CD1	325	S49 S67 T118 S156 S193 S87 S88 S137 S163 T178 S291	N5 N65 N191	Transmembrane domain: L143-S163; S203-V228 7 transmembrane receptor (rhodopsin family) domain: G41-Y290 G-protein coupled receptors signature: K90-P129; T282-K298 G-protein coupled receptors signature: L103-A147 Olfactory receptor signature: V59-L80; F177-N191; F238-G253; F274-L285; S291-W305 RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L166-L245 G-PROTEIN COUPLED RECEPTORS DM00013 A57069 15-304: F17-G306 G-protein coupled receptors motif: T110-I126	BLAST- PRODOR BLIMPS- BLOCKS PROFILES SCAN BLAST- PRODOR BLAST- PRODOR MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21	7472172CD1	313	T6 S65 S263 T76 S288	N3 N40 N63 N262	Signal peptide: M1-T36 Transmembrane domains: I10-T32; M96-M116; L191-L208 7 transmembrane receptor (rhodopsin family) domain: G39-Y287 G-protein coupled receptors signature: K88-P127; T279-Q295 G-protein coupled receptors signature: Y100-L144 Rhodopsin-like GPCR superfamily signature: S24-W48; M57-K78; Y102-I124; A138-I159; N196-I219; K269-Q295 Olfactory receptor signature: M57-K78; F174-N188; F235-G250; A271-L282; S288-V302 RECEPTOR OLFATORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L164-I242 G-PROTEIN COUPLED RECEPTORS DM00013 P23267 20-309: F15-V302 G-protein coupled receptors motif: V108-I124	SPSCAN HMMER HMMER-PFAM BLOCKS-BLIMPS PROFILESCAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODROM BLAST-DOMO MOTIFS

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
22	7472033CB1	1413	1-1413	GNN.g5902227_030.edit	1	1413
23	7472034CB1	1076	212-1076	GNN.g6087993_008.edit	1	1076
24	7472035CB1	948	1-87, 908-948, 377-798	GNN.g6088009_006	1	948
25	7472036CB1	945	1-36	GNN.g6088009_016	1	945
26	7472037CB1	966	1-46, 931-966	GNN.g6094563_010	1	966
27	7472039CB1	996	1-94, 583-996, 454-524	GNN.g6094604_016	1	996
28	7472040CB1	1014	1-843	GNN.g6165152_010	1	1014
29	4250893CB1	5122	3275-3357, 1-1445, 4869-5122, 2954-3232, 4873-4944, 4411-4479, 1740-2862, 3840-4064	SBSA02572V1 5080262H1 (LNODNOT11) 4882636F6 (LUNLTMT01) 639691X12F1 (BRSTNOT03) 2654889F6 (THYMNOT04) G4589483_CD 2831336F6 (TYMNOT03) SAFB00488F1 1559811H1 (SPLNNT04) 3345781H1 (SPLNNT09) 4220888H1 (PANCNOT07) 5090301F6 (UTRSTMR01) 504356R6 (TMLR3DT02) SBSA00168V1 SAFC01343F1 4882636T6 (LUNLTMT01)	4180 2345 508 2904 2372 339 1 3464 3448 1626 3883 1659 4586 4155 3022 621	4818 2606 909 3418 2912 4294 513 4111 3642 1884 4171 2322 5122 4681 3565 1150

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
30	6726656CB1	1241	1-311	GNN.g6524208_008	312	1241
31	7472062CB1	1155	1-211, 649-921, 1104-1155	6726656H1 (COLITUT02) GNN.g6009916_000010_002	1 1	580 1155
32	7472067CB1	1260	1-1260	GNN.g6013566_000014_004	1	1260
33	7472072CB1	945	920-945	GNN.g6165017_000177_002	1	945
34	7472074CB1	765	538-765	GNN.g6165058_000059_002	1	765
35	7472077CB1	1089	897-944	GNN.g5815499_006	1	1089
36	7472082CB1	1334	1-181	GNN.g6521401_012	1	1113
				95754986	883	1334
37	7472128CB1	960	1-22, 477-642, 940-960	GNN.g6451812_008.edit	1	960
38	7472134CB1	939	1-223, 756-804, 587-627	GNN.g6479069_014	1	939
39	7472136CB1	968	1-968	GNN.g6498052_008.edit	1	968
40	7472142CB1	1000	1-82, 975-1000, 563-684	GNN.g6524207_010.edit	1	1000
41	7472171CB1	1008	1-33, 931-1008	GNN.g6562243_020.edit	1	1008
42	7472172CB1	972	1-29, 605-972	GNN.g6525268_002.edit	1	972

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
29	4250893CB1	SYNORAT05
30	6726656CB1	COLITUT02

Table 6

Library	Vector	Library Description
COLITUT02	pINCY	Library was constructed using RNA isolated from colon tumor tissue of the ileocecal valve removed from a 29-year-old female. Pathology indicated malignant lymphoma, small cell, non-cleaved (Burkitt's lymphoma, B-cell phenotype), forming a polypoid mass in the region of the ileocecal valve, associated with intussusception and obstruction clinically. The liver and multiple (3 of 12) ileocecal region lymph nodes were also involved by lymphoma.
SYNORAT05	PSPORT1	Library was constructed using RNA isolated from the knee synovial tissue of a 62-year-old female with rheumatoid arthritis.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLOCKS IMPROVED Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1998) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score \geq GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
 - 5 a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,
 - b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,
 - c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and
 - 10 d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-21.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
- 20 5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:22-42.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
- 30 9. A method for producing a polypeptide of claim 1, the method comprising:
 - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
 - 35 b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

- 5 a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,
- b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,
- c) a polynucleotide sequence complementary to a),
- d) a polynucleotide sequence complementary to b), and
- 10 e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

15 13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization
- 20 complex is formed between said probe and said target polynucleotide or fragments thereof, and
- b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

25

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- 30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

35

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

18. A method for treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional GCREC, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- 5 a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in
- 10 the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method

15 comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of
- 20 the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at
- 25 least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- c) quantifying the amount of hybridization complex; and
- 30 d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

<110> INCYTE GENOMICS, INC.
 BAUGHN, Mariah R.
 AU-YOUNG, Janice
 YUE, Henry

<120> G-PROTEIN COUPLED RECEPTORS

<130> PI-0032 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/180,093; 60/182,045

<151> 2000-02-02; 2000-02-11

<160> 42

<170> PERL Program

<210> 1

<211> 470

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472033CD1

<220>

<221> unsure

<222> 59, 128, 309

<223> unknown or other

<400> 1

```

Met Asn Gln Thr Glu Pro Ala Gln Leu Ala Asp Gly Glu His Leu
  1      5      10
Ser Gly Tyr Ala Ser Ser Ser Asn Ser Val Arg Tyr Leu Asp Asp
      20      25      30
Arg His Pro Leu Asp Tyr Leu Asp Leu Gly Thr Val His Ala Leu
      35      40      45
Asn Thr Thr Ala Ile Asn Thr Ser Asp Leu Asn Glu Thr Xaa Ser
      50      55      60
Arg Pro Leu Asp Pro Val Leu Ile Asp Arg Phe Leu Ser Asn Arg
      65      70      75
Ala Val Asp Ser Pro Trp Tyr His Met Leu Ile Ser Met Tyr Gly
      80      85      90
Val Leu Ile Val Phe Gly Ala Leu Gly Asn Thr Leu Gly Cys Tyr
      95     100     105
Ser Pro Ser Ser Gly Ser Pro Ser Cys Ala Leu Leu Ala Ile Trp
     110     115     120
Phe Ile Leu Asn Leu Ala Ile Xaa Gly Gln Ser Lys Cys Glu Ser
     125     130     135
His Pro Ser Gly Leu Ser Asp Leu Leu Leu Cys Leu Val Thr Met
     140     145     150
Pro Leu Thr Leu Met Glu Ile Leu Ser Lys Tyr Trp Pro Tyr Gly
     155     160     165
Ser Cys Ser Ile Leu Cys Lys Thr Ile Ala Met Leu Gln Ala Leu
     170     175     180
Cys Ile Phe Val Ser Thr Ile Ser Ile Thr Ala Ile Ala Phe Asp
     185     190     195
Arg Tyr Gln Val Ile Val Tyr Pro Thr Arg Asp Ser Leu Gln Phe
     200     205     210
Val Gly Ala Val Thr Ile Leu Ala Gly Ile Trp Ala Leu Ala Leu
     215     220     225
Leu Leu Ala Ser Pro Leu Phe Val Tyr Lys Glu Leu Ile Asn Thr
     230     235     240
Asp Thr Pro Ala Leu Leu Gln Gln Ile Gly Leu Gln Asp Thr Ile

```


Pro Tyr Cys Ile	245	Arg Asn Gly Arg Phe	255
Glu Asp Trp Pro Ser	260	Tyr	270
Tyr Ser Ile Phe	275	Leu Val Pro Ile	285
Ser Leu Cys Val Gln	290	Leu	300
Ile Val Ser Val	305	Tyr Asn Lys Leu Lys	315
Ala Tyr Phe Gly Ile	320	Arg Lys Val Glu Arg	330
Val Ala Val Gln Xaa	335	Leu Ile Ser Ile Ala	345
Arg Ile Thr Val	350	Asp Phe Phe Asn Leu	360
Val Ala Val Gln Xaa	365	Tyr	375
Arg Arg Met Lys	380	Gln Ser Met Leu Val	390
Arg Thr Asn Cys Leu	395	Arg	405
Ile Phe Gly Val	410	Ser Ser Ala Cys Ser	420
Ser Trp Leu Pro Leu	425	Asn Phe Arg Lys Glu	435
Asp Phe Phe Asn Leu	440	Ile	450
Ala Asp Met Glu	455	Asn Val Ala Leu	465
Arg Ser Pro Val Thr	470	Arg	
Tyr Ala Ile Cys		Arg	
His Met Ile Gly Met		Arg	
Pro Leu Leu Tyr		Arg	
Gly Trp Leu Asn Asp		Arg	
Gln Glu Leu Leu		Arg	
Cys Arg Cys Ser Asp		Arg	
Gly His Thr Thr		Arg	
Gly Cys Asn Val Gln		Arg	
Arg Lys Tyr Gly		Arg	
Arg Arg Ile Leu Gln		Arg	
Gly Ala Gly Gly		Arg	
Ala Arg Ala Val Pro		Arg	
Gly Gly His Arg		Arg	
Leu His Asp Arg His		Arg	
Asn Ile Val His		Arg	

<210> 2

<211> 326

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472034CD1

<400> 2

Met Asp Pro Asn Gln	5	Asp Glu Ile Ser	10	Glu Leu Pro Glu Lys	15
Phe Arg Arg Ser	20	Ile Ile Lys Leu Ile	25	Lys Glu Ala Pro Glu	30
Gly Ile Pro Gly	35	Leu Glu Glu Ser Gln	40	His Trp Ile Ala Leu	45
Leu Gly Ile Leu	50	Tyr Leu Leu Ala Leu	55	Val Gly Asn Val Thr	60
Leu Phe Ile Ile	65	Trp Met Asp Pro Ser	70	Leu His Gln Ser Met	75
Leu Phe Leu Ser	80	Met Leu Ala Ala Ile	85	Asp Leu Val Leu Ala	90
Ser Thr Ala Pro	95	Lys Ala Leu Ala Val	100	Leu Leu Val His Ala	105
Glu Ile Gly Tyr	110	Ile Val Cys Leu Ile	115	Gln Met Phe Phe Ile	120
Ala Phe Ser Ser	125	Met Glu Ser Gly Val	130	Leu Val Ala Met Ala	135
Asp Arg Tyr Val	140	Ala Ile Cys His Pro	145	Leu His His Ser Thr	150
Leu His Pro Gly	155	Val Ile Gly Arg Ile	160	Gly Met Val Val Leu	165
Arg Gly Leu Leu	170	Leu Leu Ile Pro Phe	175	Pro Ile Leu Leu Gly	180
Leu Ile Phe Cys		Gln Ala Thr Ile Ile		Gly His Ala Tyr Cys	

His Met Ala Val	185	190	195
Val Lys Leu Ala Cys Ser Glu Thr Thr Val Asn			
200	205	210	
Arg Ala Tyr Gly Leu Thr Met Ala Leu Leu Val Ile Gly Leu Asp			
215	220	225	
Val Leu Ala Ile Gly Val Ser Tyr Ala His Ile Leu Gln Ala Val			
230	235	240	
Leu Lys Val Pro Gly Ser Glu Ala Arg Leu Lys Ala Phe Ser Thr			
245	250	255	
Cys Gly Ser His Ile Cys Val Ile Leu Val Phe Tyr Val Pro Gly			
260	265	270	
Ile Phe Ser Phe Leu Thr His Arg Phe Gly His His Val Pro His			
275	280	285	
His Val His Val Leu Leu Ala Thr Arg Tyr Leu Leu Met Pro Pro			
290	295	300	
Ala Leu Asn Pro Leu Val Tyr Gly Val Lys Thr Gln Gln Ile Arg			
305	310	315	
Gln Arg Val Leu Arg Val Phe Thr Gln Lys Asp			
320	325		

<210> 3

<211> 315

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472035CD1

<400> 3

Met Glu Thr Pro Ala Ser Phe Leu Leu Val Gly Ile Pro Gly Leu	
1 5 10 15	
Gln Ser Ser His Leu Trp Leu Ala Ile Ser Leu Ser Ala Met Tyr	
20 25 30	
Ile Ile Ala Leu Leu Gly Asn Thr Ile Ile Val Thr Ala Ile Trp	
35 40 45	
Met Asp Ser Thr Arg His Glu Pro Met Tyr Cys Phe Leu Cys Val	
50 55 60	
Leu Ala Ala Val Asp Ile Val Met Ala Ser Ser Val Val Pro Lys	
65 70 75	
Met Val Ser Ile Phe Cys Ser Gly Asp Ser Ser Ile Ser Phe Ser	
80 85 90	
Ala Cys Phe Thr Gln Met Phe Phe Val His Leu Ala Thr Ala Val	
95 100 105	
Glu Thr Gly Leu Leu Leu Thr Met Ala Phe Asp Arg Tyr Val Ala	
110 115 120	
Ile Cys Lys Pro Leu His Tyr Lys Arg Ile Leu Thr Pro Gln Val	
125 130 135	
Met Leu Gly Met Ser Met Ala Ile Thr Ile Arg Ala Ile Ile Ala	
140 145 150	
Ile Thr Pro Leu Ser Trp Met Val Ser His Leu Pro Phe Cys Gly	
155 160 165	
Ser Asn Val Val Val His Ser Tyr Cys Glu His Ile Ala Leu Ala	
170 175 180	
Arg Leu Ala Cys Ala Asp Pro Val Pro Ser Ser Leu Tyr Ser Leu	
185 190 195	
Ile Gly Ser Ser Leu Met Val Gly Ser Asp Val Ala Phe Ile Ala	
200 205 210	
Ala Ser Tyr Ile Leu Ile Leu Lys Ala Val Phe Gly Leu Ser Ser	
215 220 225	
Lys Thr Ala Gln Leu Lys Ala Leu Ser Thr Cys Gly Ser His Val	
230 235 240	
Gly Val Met Ala Leu Tyr Tyr Leu Pro Gly Met Ala Ser Ile Tyr	
245 250 255	
Ala Ala Trp Leu Gly Gln Asp Val Val Pro Leu His Thr Gln Val	
260 265 270	
Leu Leu Ala Asp Leu Tyr Val Ile Ile Pro Ala Thr Leu Asn Pro	

	275		280	285
Ile Ile Tyr Gly Met	Arg Thr Lys Gln	Leu Arg Glu Arg Ile	Trp	
	290	295	300	
Ser Tyr Leu Met His	Val Leu Phe Asp	His Ser Asn Leu Gly	Ser	
	305	310	315	

<210> 4
 <211> 314
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472036CD1

<400> 4
 Met Ser Ala Ser Asn Ile Thr Leu Thr His Pro Thr Ala Phe Leu
 1 5 10 15
 Leu Val Gly Ile Pro Gly Leu Glu His Leu His Ile Trp Ile Ser
 20 25 30
 Ile Pro Phe Cys Leu Ala Tyr Thr Leu Ala Leu Leu Gly Asn Cys
 35 40 45
 Thr Leu Leu Leu Ile Ile Gln Ala Asp Ala Ala Leu His Glu Pro
 50 55 60
 Met Tyr Leu Phe Leu Ala Met Leu Ala Ala Ile Asp Leu Val Leu
 65 70 75
 Ser Ser Ser Ala Leu Pro Lys Met Leu Ala Ile Phe Trp Phe Arg
 80 85 90
 Asp Arg Glu Ile Asn Phe Phe Ala Cys Leu Ala Gln Met Phe Phe
 95 100 105
 Leu His Ser Phe Ser Ile Met Glu Ser Ala Val Leu Leu Ala Met
 110 115 120
 Ala Phe Asp Arg Tyr Val Ala Ile Cys Lys Pro Leu His Tyr Thr
 125 130 135
 Lys Val Leu Thr Gly Ser Leu Ile Thr Lys Ile Gly Met Ala Ala
 140 145 150
 Val Ala Arg Ala Val Thr Leu Met Thr Pro Leu Pro Phe Leu Leu
 155 160 165
 Arg Cys Phe His Tyr Cys Arg Gly Pro Val Ile Ala His Cys Tyr
 170 175 180
 Cys Glu His Met Ala Val Val Arg Leu Ala Cys Gly Asp Thr Ser
 185 190 195
 Phe Asn Asn Ile Tyr Gly Ile Ala Val Ala Met Phe Ile Val Val
 200 205 210
 Leu Asp Leu Leu Leu Val Ile Leu Ser Tyr Ile Phe Ile Leu Gln
 215 220 225
 Ala Val Leu Leu Leu Ala Ser Gln Glu Ala Arg Tyr Lys Ala Phe
 230 235 240
 Gly Thr Cys Val Ser His Ile Gly Ala Ile Leu Ala Phe Tyr Thr
 245 250 255
 Thr Val Val Ile Ser Ser Val Met His Arg Val Ala Arg His Ala
 260 265 270
 Ala Pro His Val His Ile Leu Leu Ala Asn Phe Tyr Leu Leu Phe
 275 280 285
 Pro Pro Met Val Asn Pro Ile Ile Tyr Gly Val Lys Thr Lys Gln
 290 295 300
 Ile Arg Glu Ser Ile Leu Gly Val Phe Pro Arg Lys Asp Met
 305 310

<210> 5
 <211> 321
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature

<223> Incyte ID No: 7472037CD1

<400> 5

```

Met Ala His Gln Ala Pro Glu Lys Gln Gln Asp Asn Gly Thr Trp
 1      5      10      15
Leu Val Thr Glu Phe Leu Leu Val Gly Phe Ser Asn Leu Pro Glu
 20      25      30
Leu Arg Pro Thr Leu Phe Ile Leu Phe Leu Leu Thr Tyr Leu Val
 35      40      45
Thr Leu Ser Gly Asn Ala Thr Ile Ile Thr Ile Ile Gln Val Asp
 50      55      60
Arg Thr Leu His Thr Pro Met Tyr Arg Phe Leu Ala Val Leu Ser
 65      70      75
Leu Ser Glu Thr Cys Tyr Thr Leu Val Thr Ile Pro Asn Met Leu
 80      85      90
Ala His Leu Leu Met Glu Ser Gln Ala Ile Ser Ile Ala Gly Cys
 95      100     105
Arg Ala Gln Met Phe Phe Phe Leu Gly Leu Gly Cys Ser His Cys
110      115     120
Phe Leu Leu Thr Leu Met Gly Tyr Asp Arg Tyr Val Ala Ile Cys
125      130     135
His Pro Leu Arg Tyr Ser Val Ile Met Arg Pro Thr Val Cys Leu
140      145     150
Cys Leu Gly Ala Leu Val Phe Cys Ser Gly Phe Ser Val Ala Leu
155      160     165
Ile Glu Thr Cys Met Ile Phe Ser Ser Pro Phe Cys Gly Ala Gly
170      175     180
His Val Glu His Phe Phe Cys Asp Ile Ala Pro Val Leu Lys Leu
185      190     195
Ser Cys Asp Glu Ser Ser Leu Lys Gly Leu Gly Ile Phe Phe Leu
200      205     210
Ser Ile Leu Val Val Leu Val Ser Phe Leu Phe Ile Leu Leu Ser
215      220     225
Tyr Ala Phe Ile Val Ala Ala Ile Val Arg Ile Pro Ser Ala Ser
230      235     240
Gly Arg Arg Lys Ala Phe Ser Thr Cys Ala Ala His Leu Thr Val
245      250     255
Val Ile Val His Phe Gly Cys Ala Ser Ile Ile Tyr Leu Arg Pro
260      265     270
Asp Ser Gly Ala Asn Pro Ser Gln Asp Arg Leu Val Ala Val Phe
275      280     285
Tyr Thr Val Val Thr Pro Leu Leu Asn Pro Val Val Tyr Thr Leu
290      295     300
Arg Asn Lys Glu Val Arg Val Ala Leu Arg Lys Asn Leu Ala Arg
305      310     315
Gly Cys Gly Ala Phe Lys
320

```

<210> 6

<211> 331

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472039CD1

<400> 6

```

Met Ser Pro Asp Gly Asn His Ser Ser Asp Pro Thr Glu Phe Val
 1      5      10      15
Leu Ala Gly Leu Pro Asn Leu Asn Ser Ala Arg Val Glu Leu Phe
 20      25      30
Ser Val Phe Leu Leu Val Tyr Leu Leu Asn Leu Thr Gly Asn Val
 35      40      45
Leu Ile Val Gly Val Val Arg Ala Asp Thr Arg Leu Gln Thr Pro
 50      55      60
Met Tyr Phe Phe Leu Gly Asn Leu Ser Cys Leu Glu Ile Leu Leu

```

				65					70					75
Thr	Ser	Val	Ile	Ile	Pro	Lys	Met	Leu	Ser	Asn	Phe	Leu	Ser	Arg
				80										90
Gln	His	Thr	Ile	Ser	Phe	Ala	Ala	Cys	Ile	Thr	Gln	Phe	Tyr	Phe
				95										105
Tyr	Phe	Phe	Leu	Gly	Ala	Ser	Glu	Phe	Leu	Leu	Leu	Ala	Val	Met
				110										120
Ser	Ala	Asp	Arg	Tyr	Leu	Ala	Ile	Cys	His	Pro	Leu	Arg	Tyr	Pro
				125										135
Leu	Leu	Met	Ser	Gly	Ala	Val	Cys	Phe	Arg	Val	Ala	Leu	Ala	Cys
				140										150
Trp	Val	Gly	Gly	Leu	Val	Pro	Val	Leu	Gly	Pro	Thr	Val	Ala	Val
				155										165
Ala	Leu	Leu	Pro	Phe	Cys	Lys	Gln	Gly	Ala	Val	Val	Gln	His	Phe
				170										180
Phe	Cys	Asp	Ser	Gly	Pro	Leu	Leu	Arg	Leu	Ala	Cys	Thr	Asn	Thr
				185										195
Lys	Lys	Leu	Glu	Glu	Thr	Asp	Phe	Val	Leu	Ala	Ser	Leu	Val	Ile
				200										210
Val	Ser	Ser	Leu	Leu	Ile	Thr	Ala	Val	Ser	Tyr	Gly	Leu	Ile	Val
				215										225
Leu	Ala	Val	Leu	Ser	Ile	Pro	Ser	Ala	Ser	Gly	Arg	Gln	Lys	Ala
				230										240
Phe	Ser	Thr	Cys	Thr	Ser	His	Leu	Ile	Val	Val	Thr	Leu	Phe	Tyr
				245										255
Gly	Ser	Ala	Ile	Phe	Leu	Tyr	Val	Arg	Pro	Ser	Gln	Ser	Gly	Ser
				260										270
Val	Asp	Thr	Asn	Trp	Ala	Val	Thr	Val	Ile	Thr	Thr	Phe	Val	Thr
				275										285
Pro	Leu	Leu	Asn	Pro	Phe	Ile	Tyr	Ala	Leu	Arg	Asn	Glu	Gln	Val
				290										300
Lys	Glu	Ala	Leu	Lys	Asp	Met	Phe	Arg	Lys	Val	Val	Ala	Gly	Val
				305										315
Leu	Gly	Asn	Leu	Leu	Leu	Asp	Lys	Cys	Leu	Ser	Glu	Lys	Ala	Val
				320										330
Lys														

<210> 7
 <211> 337
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472040CD1

Met	Gly	Asn	Asp	Ser	Val	Ser	Tyr	Glu	Tyr	Gly	Asp	Tyr	Ser	Asp
1				5					10					15
Leu	Ser	Asp	Arg	Pro	Val	Asp	Cys	Leu	Asp	Gly	Ala	Cys	Leu	Ala
				20					25					30
Ile	Asp	Pro	Leu	Arg	Val	Ala	Pro	Leu	Pro	Leu	Tyr	Ala	Ala	Ile
				35					40					45
Phe	Leu	Val	Gly	Val	Pro	Gly	Asn	Ala	Met	Val	Ala	Trp	Val	Ala
				50					55					60
Gly	Lys	Val	Ala	Arg	Arg	Arg	Val	Gly	Ala	Thr	Trp	Leu	Leu	His
				65					70					75
Leu	Ala	Val	Ala	Asp	Leu	Leu	Cys	Cys	Leu	Ser	Leu	Pro	Ile	Leu
				80					85					90
Ala	Val	Pro	Ile	Ala	Arg	Gly	Gly	His	Trp	Pro	Tyr	Gly	Ala	Val
				95					100					105
Gly	Cys	Arg	Ala	Leu	Pro	Ser	Ile	Ile	Leu	Leu	Thr	Met	Tyr	Ala
				110					115					120
Ser	Val	Leu	Leu	Leu	Ala	Ala	Leu	Ser	Ala	Asp	Leu	Cys	Phe	Leu
				125					130					135
Ala	Leu	Gly	Pro	Ala	Trp	Trp	Ser	Thr	Val	Gln	Arg	Ala	Cys	Gly

				140					145					150
Val	Gln	Val	Ala	Cys	Gly	Ala	Ala	Trp	Thr	Leu	Ala	Leu	Leu	Leu
				155					160					165
Thr	Val	Pro	Ser	Ala	Ile	Tyr	Arg	Arg	Leu	His	Gln	Glu	His	Phe
				170					175					180
Pro	Ala	Arg	Leu	Gln	Cys	Val	Val	Asp	Tyr	Gly	Gly	Ser	Ser	Ser
				185					190					195
Thr	Glu	Asn	Ala	Val	Thr	Ala	Ile	Arg	Phe	Leu	Phe	Gly	Phe	Leu
				200					205					210
Gly	Pro	Leu	Val	Ala	Val	Ala	Ser	Cys	His	Ser	Ala	Leu	Leu	Cys
				215					220					225
Trp	Ala	Ala	Arg	Arg	Cys	Arg	Pro	Leu	Gly	Thr	Ala	Ile	Val	Val
				230					235					240
Gly	Phe	Phe	Val	Cys	Trp	Ala	Pro	Tyr	His	Leu	Leu	Gly	Leu	Val
				245					250					255
Leu	Thr	Val	Ala	Ala	Pro	Asn	Ser	Ala	Leu	Leu	Ala	Arg	Ala	Leu
				260					265					270
Arg	Ala	Glu	Pro	Leu	Ile	Val	Gly	Leu	Ala	Leu	Ala	His	Ser	Cys
				275					280					285
Leu	Asn	Pro	Met	Leu	Phe	Leu	Tyr	Phe	Gly	Arg	Ala	Gln	Leu	Arg
				290					295					300
Arg	Ser	Leu	Pro	Ala	Ala	Cys	His	Trp	Ala	Leu	Arg	Glu	Ser	Gln
				305					310					315
Gly	Gln	Asp	Glu	Ser	Val	Asp	Ser	Lys	Lys	Ser	Thr	Ser	His	Asp
				320					325					330
Leu	Val	Ser	Glu	Met	Glu	Val								
				335										

<210> 8
 <211> 1473
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4250893CD1

<400>	8													
Met	Ala	Gly	Gly	Ala	Trp	Gly	Arg	Leu	Ala	Cys	Tyr	Leu	Glu	Phe
1				5					10					15
Leu	Lys	Lys	Glu	Glu	Leu	Lys	Glu	Phe	Gln	Leu	Leu	Leu	Ala	Asn
				20					25					30
Lys	Ala	His	Ser	Arg	Ser	Ser	Ser	Gly	Glu	Thr	Pro	Ala	Gln	Pro
				35					40					45
Glu	Lys	Thr	Ser	Gly	Met	Glu	Val	Ala	Ser	Tyr	Leu	Val	Ala	Gln
				50					55					60
Tyr	Gly	Glu	Gln	Arg	Ala	Trp	Asp	Leu	Ala	Leu	His	Thr	Trp	Glu
				65					70					75
Gln	Met	Gly	Leu	Arg	Ser	Leu	Cys	Ala	Gln	Ala	Gln	Glu	Gly	Ala
				80					85					90
Gly	His	Ser	Pro	Ser	Phe	Pro	Tyr	Ser	Pro	Ser	Glu	Pro	His	Leu
				95					100					105
Gly	Ser	Pro	Ser	Gln	Pro	Thr	Ser	Thr	Ala	Val	Leu	Met	Pro	Trp
				110					115					120
Ile	His	Glu	Leu	Pro	Ala	Gly	Cys	Thr	Gln	Gly	Ser	Glu	Arg	Arg
				125					130					135
Val	Leu	Arg	Gln	Leu	Pro	Asp	Thr	Ser	Gly	Arg	Arg	Trp	Arg	Glu
				140					145					150
Ile	Ser	Ala	Ser	His	Val	Tyr	Gln	Ala	Leu	Pro	Ser	Ser	Pro	Asp
				155					160					165
His	Glu	Ser	Pro	Ser	Gln	Glu	Ser	Pro	Asn	Ala	Pro	Thr	Ser	Thr
				170					175					180
Ala	Val	Leu	Gly	Ser	Trp	Gly	Ser	Pro	Pro	Gln	Pro	Ser	Leu	Ala
				185					190					195
Pro	Arg	Glu	Gln	Glu	Ala	Pro	Gly	Thr	Gln	Trp	Pro	Leu	Asp	Glu
				200					205					210
Thr	Ser	Gly	Ile	Tyr	Tyr	Thr	Glu	Ile	Arg	Glu	Arg	Glu	Arg	Glu

				215					220					225
Lys	Ser	Glu	Lys	Gly	Arg	Pro	Pro	Trp	Ala	Ala	Val	Val	Gly	Thr
				230					235					240
Pro	Pro	Gln	Ala	His	Thr	Ser	Leu	Gln	Pro	His	His	His	Pro	Trp
				245					250					255
Glu	Pro	Ser	Val	Arg	Glu	Ser	Leu	Cys	Ser	Thr	Trp	Pro	Trp	Lys
				260					265					270
Asn	Glu	Asp	Phe	Asn	Gln	Lys	Phe	Thr	Gln	Leu	Leu	Leu	Leu	Gln
				275					280					285
Arg	Pro	His	Pro	Arg	Ser	Gln	Asp	Pro	Leu	Val	Lys	Arg	Ser	Trp
				290					295					300
Pro	Asp	Tyr	Val	Glu	Glu	Asn	Arg	Gly	His	Leu	Ile	Glu	Ile	Arg
				305					310					315
Asp	Leu	Phe	Gly	Pro	Gly	Leu	Asp	Thr	Gln	Glu	Pro	Arg	Ile	Val
				320					325					330
Ile	Leu	Gln	Gly	Ala	Ala	Gly	Ile	Gly	Lys	Ser	Thr	Leu	Ala	Arg
				335					340					345
Gln	Val	Lys	Glu	Ala	Trp	Gly	Arg	Gly	Gln	Leu	Tyr	Gly	Asp	Arg
				350					355					360
Phe	Gln	His	Val	Phe	Tyr	Phe	Ser	Cys	Arg	Glu	Leu	Ala	Gln	Ser
				365					370					375
Lys	Val	Val	Ser	Leu	Ala	Glu	Leu	Ile	Gly	Lys	Asp	Gly	Thr	Ala
				380					385					390
Thr	Pro	Ala	Pro	Ile	Arg	Gln	Ile	Leu	Ser	Arg	Pro	Glu	Arg	Leu
				395					400					405
Leu	Phe	Ile	Leu	Asp	Gly	Val	Asp	Glu	Pro	Gly	Trp	Val	Leu	Gln
				410					415					420
Glu	Pro	Ser	Ser	Glu	Leu	Cys	Leu	His	Trp	Ser	Gln	Pro	Gln	Pro
				425					430					435
Ala	Asp	Ala	Leu	Leu	Gly	Ser	Leu	Leu	Gly	Lys	Thr	Ile	Leu	Pro
				440					445					450
Glu	Ala	Ser	Phe	Leu	Ile	Thr	Ala	Arg	Thr	Thr	Ala	Leu	Gln	Asn
				455					460					465
Leu	Ile	Pro	Ser	Leu	Glu	Gln	Ala	Arg	Trp	Val	Glu	Val	Leu	Gly
				470					475					480
Phe	Ser	Glu	Ser	Ser	Arg	Lys	Glu	Tyr	Phe	Tyr	Arg	Tyr	Phe	Thr
				485					490					495
Asp	Glu	Arg	Gln	Ala	Ile	Arg	Ala	Phe	Arg	Leu	Val	Lys	Ser	Asn
				500					505					510
Lys	Glu	Leu	Trp	Ala	Leu	Cys	Leu	Val	Pro	Trp	Val	Ser	Trp	Leu
				515					520					525
Ala	Cys	Thr	Cys	Leu	Met	Gln	Gln	Met	Lys	Arg	Lys	Glu	Lys	Leu
				530					535					540
Thr	Leu	Thr	Ser	Lys	Thr	Thr	Thr	Thr	Leu	Cys	Leu	His	Tyr	Leu
				545					550					555
Ala	Gln	Ala	Leu	Gln	Ala	Gln	Pro	Leu	Gly	Pro	Gln	Leu	Arg	Asp
				560					565					570
Leu	Cys	Ser	Leu	Ala	Ala	Glu	Gly	Ile	Trp	Gln	Lys	Lys	Thr	Leu
				575					580					585
Phe	Ser	Pro	Asp	Asp	Leu	Arg	Lys	His	Gly	Leu	Asp	Gly	Ala	Ile
				590					595					600
Ile	Ser	Thr	Phe	Leu	Lys	Met	Gly	Ile	Leu	Gln	Glu	His	Pro	Ile
				605					610					615
Pro	Leu	Ser	Tyr	Ser	Phe	Ile	His	Leu	Cys	Phe	Gln	Glu	Phe	Phe
				620					625					630
Ala	Ala	Met	Ser	Tyr	Val	Leu	Glu	Asp	Glu	Lys	Gly	Arg	Gly	Lys
				635					640					645
His	Ser	Asn	Cys	Ile	Ile	Asp	Leu	Glu	Lys	Thr	Leu	Glu	Ala	Tyr
				650					655					660
Gly	Ile	His	Gly	Leu	Phe	Gly	Ala	Ser	Thr	Thr	Arg	Phe	Leu	Leu
				665					670					675
Gly	Leu	Leu	Ser	Asp	Glu	Gly	Glu	Arg	Glu	Met	Glu	Asn	Ile	Phe
				680					685					690
His	Cys	Arg	Leu	Ser	Gln	Gly	Arg	Asn	Leu	Met	Gln	Trp	Val	Pro
				695					700					705
Ser	Leu	Gln	Leu	Leu	Leu	Gln	Pro	His	Ser	Leu	Glu	Ser	Leu	His
				710					715					720

Cys	Leu	Tyr	Glu	Thr	Arg	Asn	Lys	Thr	Phe	Leu	Thr	Gln	Val	Met
				725					730					735
Ala	His	Phe	Glu	Glu	Met	Gly	Met	Cys	Val	Glu	Thr	Asp	Met	Glu
				740					745					750
Leu	Leu	Val	Cys	Thr	Phe	Cys	Ile	Lys	Phe	Ser	Arg	His	Val	Lys
				755					760					765
Lys	Leu	Gln	Leu	Ile	Glu	Gly	Arg	Gln	His	Arg	Ser	Thr	Trp	Ser
				770					775					780
Pro	Thr	Met	Val	Val	Leu	Phe	Arg	Trp	Val	Pro	Val	Thr	Asp	Ala
				785					790					795
Tyr	Trp	Gln	Ile	Leu	Phe	Ser	Val	Leu	Lys	Val	Thr	Arg	Asn	Leu
				800					805					810
Lys	Glu	Leu	Asp	Leu	Ser	Gly	Asn	Ser	Leu	Ser	His	Ser	Ala	Val
				815					820					825
Lys	Ser	Leu	Cys	Lys	Thr	Leu	Arg	Arg	Pro	Arg	Cys	Leu	Leu	Glu
				830					835					840
Thr	Leu	Arg	Leu	Ala	Gly	Cys	Gly	Leu	Thr	Ala	Glu	Asp	Cys	Lys
				845					850					855
Asp	Leu	Ala	Phe	Gly	Leu	Arg	Ala	Asn	Gln	Thr	Leu	Thr	Glu	Leu
				860					865					870
Asp	Leu	Ser	Phe	Asn	Val	Leu	Thr	Asp	Ala	Gly	Ala	Lys	His	Leu
				875					880					885
Cys	Gln	Arg	Leu	Arg	Gln	Pro	Ser	Cys	Lys	Leu	Gln	Arg	Leu	Gln
				890					895					900
Leu	Val	Ser	Cys	Gly	Leu	Thr	Ser	Asp	Cys	Cys	Gln	Asp	Leu	Ala
				905					910					915
Ser	Val	Leu	Ser	Ala	Ser	Pro	Ser	Leu	Lys	Glu	Leu	Asp	Leu	Gln
				920					925					930
Gln	Asn	Asn	Leu	Asp	Asp	Val	Gly	Val	Arg	Leu	Leu	Cys	Glu	Gly
				935					940					945
Leu	Arg	His	Pro	Ala	Cys	Lys	Leu	Ile	Arg	Leu	Gly	Leu	Asp	Gln
				950					955					960
Thr	Thr	Leu	Ser	Asp	Glu	Met	Arg	Gln	Glu	Leu	Arg	Ala	Leu	Glu
				965					970					975
Gln	Glu	Lys	Pro	Gln	Leu	Leu	Ile	Phe	Ser	Arg	Arg	Lys	Pro	Ser
				980					985					990
Val	Met	Thr	Pro	Thr	Glu	Gly	Leu	Asp	Thr	Gly	Glu	Met	Ser	Asn
				995					1000					1005
Ser	Thr	Ser	Ser	Leu	Lys	Arg	Gln	Arg	Leu	Gly	Ser	Glu	Arg	Ala
				1010					1015					1020
Ala	Ser	His	Val	Ala	Gln	Ala	Asn	Leu	Lys	Leu	Leu	Asp	Val	Ser
				1025					1030					1035
Lys	Ile	Phe	Pro	Ile	Ala	Glu	Ile	Ala	Glu	Glu	Ser	Ser	Pro	Glu
				1040					1045					1050
Val	Val	Pro	Val	Glu	Leu	Leu	Cys	Val	Pro	Ser	Pro	Ala	Ser	Gln
				1055					1060					1065
Gly	Asp	Leu	His	Thr	Lys	Pro	Leu	Gly	Thr	Asp	Asp	Asp	Phe	Trp
				1070					1075					1080
Gly	Pro	Thr	Gly	Pro	Val	Ala	Thr	Glu	Val	Val	Asp	Lys	Glu	Lys
				1085					1090					1095
Asn	Leu	Tyr	Arg	Val	His	Phe	Pro	Val	Ala	Gly	Ser	Tyr	Arg	Trp
				1100					1105					1110
Pro	Asn	Thr	Gly	Leu	Cys	Phe	Val	Met	Arg	Glu	Ala	Val	Thr	Val
				1115					1120					1125
Glu	Ile	Glu	Phe	Cys	Val	Trp	Asp	Gln	Phe	Leu	Gly	Glu	Ile	Asn
				1130					1135					1140
Pro	Gln	His	Ser	Trp	Met	Val	Ala	Gly	Pro	Leu	Leu	Asp	Ile	Lys
				1145					1150					1155
Ala	Glu	Pro	Gly	Ala	Val	Glu	Ala	Val	His	Leu	Pro	His	Phe	Val
				1160					1165					1170
Ala	Leu	Gln	Gly	Gly	His	Val	Asp	Thr	Ser	Leu	Phe	Gln	Met	Ala
				1175					1180					1185
His	Phe	Lys	Glu	Glu	Gly	Met	Leu	Leu	Glu	Lys	Pro	Ala	Arg	Val
				1190					1195					1200
Glu	Leu	His	His	Ile	Val	Leu	Glu	Asn	Pro	Ser	Phe	Ser	Pro	Leu
				1205					1210					1215
Gly	Val	Leu	Leu	Lys	Met	Ile	His	Asn	Ala	Leu	Arg	Phe	Ile	Pro


```

1220      1225      1230
Val Thr Ser Val Val Leu Leu Tyr His Arg Val His Pro Glu Glu
1235      1240      1245
Val Thr Phe His Leu Tyr Leu Ile Pro Ser Asp Cys Ser Ile Arg
1250      1255      1260
Lys Ala Ile Asp Asp Leu Glu Met Lys Phe Gln Phe Val Arg Ile
1265      1270      1275
His Lys Pro Pro Pro Leu Thr Pro Leu Tyr Met Gly Cys Arg Tyr
1280      1285      1290
Thr Val Ser Gly Ser Gly Ser Gly Met Leu Glu Ile Leu Pro Lys
1295      1300      1305
Glu Leu Glu Leu Cys Tyr Arg Ser Pro Gly Glu Asp Gln Leu Phe
1310      1315      1320
Ser Glu Ser Tyr Val Gly His Leu Gly Ser Gly Ile Arg Leu Gln
1325      1330      1335
Val Lys Asp Lys Lys Asp Glu Thr Leu Val Trp Glu Ala Leu Val
1340      1345      1350
Lys Pro Gly Asp Leu Met Pro Ala Thr Thr Leu Ile Pro Pro Ala
1355      1360      1365
Arg Ile Ala Val Pro Ser Pro Leu Asp Ala Pro Gln Leu Leu His
1370      1375      1380
Phe Val Asp Gln Tyr Arg Glu Gln Leu Ile Ala Arg Val Thr Ser
1385      1390      1395
Val Glu Val Val Leu Asp Lys Leu His Gly Gln Val Leu Ser Gln
1400      1405      1410
Glu Gln Tyr Glu Arg Val Leu Ala Glu Asn Thr Arg Pro Ser Gln
1415      1420      1425
Met Arg Lys Leu Phe Ser Leu Ser Gln Ser Trp Asp Arg Lys Cys
1430      1435      1440
Lys Asp Gly Leu Tyr Gln Ala Leu Lys Glu Thr His Pro His Leu
1445      1450      1455
Ile Met Glu Leu Trp Glu Lys Gly Ser Lys Lys Gly Leu Leu Pro
1460      1465      1470
Leu Ser Ser

```

<210> 9
 <211> 328
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 6726656CD1

```

<400> 9
Met Lys Leu Trp Met Glu Ser His Leu Ile Val Pro Glu Thr Arg
1      5      10      15
Pro Ser Pro Arg Met Met Ser Asn Gln Thr Leu Val Thr Glu Phe
20      25      30
Ile Leu Gln Gly Phe Ser Glu His Pro Glu Tyr Arg Val Phe Leu
35      40      45
Phe Ser Cys Phe Leu Phe Leu Tyr Ser Gly Ala Leu Thr Gly Asn
50      55      60
Val Leu Ile Thr Leu Ala Ile Thr Phe Asn Pro Gly Leu His Ala
65      70      75
Pro Met Tyr Phe Phe Leu Leu Asn Leu Ala Thr Met Asp Ile Ile
80      85      90
Cys Thr Ser Ser Ile Met Pro Lys Ala Leu Ala Ser Leu Val Ser
95      100      105
Glu Glu Ser Ser Ile Ser Tyr Gly Gly Cys Met Ala Gln Leu Tyr
110      115      120
Phe Leu Thr Trp Ala Ala Ser Ser Glu Leu Leu Leu Leu Thr Val
125      130      135
Met Ala Tyr Asp Arg Tyr Ala Ala Ile Cys His Pro Leu His Tyr
140      145      150
Ser Ser Met Met Ser Lys Val Phe Cys Ser Gly Leu Ala Thr Ala

```

Val Trp Leu Leu	155	Ala Val Asn Thr	160	Ile His Thr Gly	165
	170		175		180
Met Leu Arg Leu	185	Phe Cys Gly Pro	190	Val Ile Ile His	195
	200		205		210
Phe Cys Glu Val	215	Pro Leu Leu Leu	220	Ser Cys Ser Ser	225
	230		235		240
Tyr Val Asn Gly	245	Val Met Ile Val Leu	250	Ala Asp Ala Phe Tyr	255
	260		265		270
Ile Val Asn Phe	275	Leu Met Thr Ile Ala	280	Ser Tyr Gly Phe Ile	285
	290		295		300
Ser Ser Ile Leu	305	Lys Val Lys Thr Ala	310	Trp Gly Arg Gln Lys	315
	320		325		
Phe Ser Thr Cys		Ser Ser His Leu Thr		Val Val Cys Met Tyr	
Thr Ala Val Phe		Tyr Ala Tyr Ile Ser		Pro Val Ser Gly Tyr	
Ala Gly Lys Ser		Lys Leu Ala Gly Leu		Leu Tyr Thr Val Leu	
Pro Thr Leu Asn		Pro Leu Ile Tyr Thr		Leu Arg Asn Lys Glu	
Lys Ala Ala Leu		Arg Lys Leu Phe Pro		Phe Phe Arg Asn	

<210> 10

<211> 384

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472062CD1

<400> 10

Met Asn Val Leu Leu	1	Ala Asp Ser Asn Ser	10	Asn Lys Lys Ile Val	15
	5		25		30
His Lys His Ile Cys	20	Ser Leu Gln Ser Ala	35	Pro Lys Thr Thr Asn	40
	35		40		45
Leu Gln Pro Ser Ile	40	Ser Asp Ile Leu Leu	55	Ser Val Glu Ser Asn	60
	50		60		65
Asp Arg Lys Asn Val	65	Ser Lys Ile Lys Gly	70	Asp Cys Phe Asn Thr	75
	80		85		90
Arg Val Ser Cys Asp	95	Ser Lys Ile Thr Ser	100	Met Glu Asn Asn Thr	105
	110		115		120
Glu Val Ser Glu Phe	125	Ile Leu Leu Gly Leu	130	Thr Asn Ala Pro Glu	135
	140		145		150
Leu Gln Val Pro Leu	155	Phe Ile Met Phe Thr	160	Leu Ile Tyr Leu Ile	165
	170		175		180
Thr Leu Thr Gly Asn	185	Leu Gly Met Ile Ile	190	Leu Ile Leu Leu Asp	195
	200		205		210
Ser His Leu His Thr	215	Pro Met Tyr Phe Phe	220	Leu Ser Asn Leu Ser	225
	230		235		240
Leu Ala Gly Ile Gly	240	Tyr Ser Ser Ala Val	245	Thr Pro Lys Val Leu	250
	255		260		265
Thr Gly Leu Leu Ile	270	Glu Asp Lys Ala Ile	275	Ser Tyr Ser Ala Cys	280
	285		290		295
Ala Ala Gln Met Phe	300	Phe Cys Ala Val Phe	305	Ala Thr Val Glu Asn	310
	315		320		325
Tyr Leu Leu Ser Ser	330	Met Ala Tyr Asp Arg	335	Tyr Ala Ala Val Cys	340
	345		350		355
Asn Pro Leu His Tyr	360	Thr Thr Thr Met Thr	365	Thr Arg Val Cys Ala	370
	375		380		385
Cys Leu Ala Ile Gly	390	Cys Tyr Val Ile Gly	395	Phe Leu Asn Ala Ser	400
	405		410		415
Ile Gln Ile Gly Asp	420	Thr Phe Arg Leu Ser	425	Phe Cys Met Ser Asn	430
	435		440		445
Val Ile His His Phe	450	Phe Cys Asp Lys Pro	455	Ala Val Ile Thr Leu	460

	245		250		255
Thr Cys Ser Glu Lys His Ile Ser Glu Leu Ile Leu Val Leu Ile	260		265		270
Ser Ser Phe Asn Val Phe Phe Ala Leu Leu Val Thr Leu Ile Ser	275		280		285
Tyr Leu Phe Ile Leu Ile Thr Ile Leu Lys Arg His Thr Gly Lys	290		295		300
Gly Tyr Gln Lys Pro Leu Ser Thr Cys Gly Ser His Leu Ile Ala	305		310		315
Ile Phe Leu Phe Tyr Ile Thr Val Ile Ile Met Tyr Ile Arg Pro	320		325		330
Ser Ser Ser His Ser Met Asp Thr Asp Lys Ile Ala Ser Val Phe	335		340		345
Tyr Thr Met Ile Ile Pro Met Leu Ser Pro Ile Val Tyr Thr Leu	350		355		360
Arg Asn Lys Asp Val Lys Asn Ala Phe Met Lys Val Val Glu Lys	365		370		375
Ala Lys Tyr Ser Leu Asp Ser Val Phe	380				

<210> 11

<211> 419

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472067CD1

<400> 11

Met Leu Ala Ala Ala Phe Ala Asp Ser Asn Ser Ser Ser Met Asn	5	10	15
Val Ser Phe Ala His Leu His Phe Ala Gly Tyr Leu Pro Ser	20	25	30
Asp Ser Gln Asp Trp Arg Thr Ile Ile Pro Ala Leu Leu Val Ala	35	40	45
Val Cys Leu Val Gly Phe Val Gly Asn Leu Cys Val Ile Gly Ile	50	55	60
Leu Leu His Asn Ala Trp Lys Gly Lys Pro Ser Met Ile His Ser	65	70	75
Leu Ile Leu Asn Leu Ser Leu Ala Asp Leu Ser Leu Leu Leu Phe	80	85	90
Ser Ala Pro Ile Arg Ala Thr Ala Tyr Ser Lys Ser Val Trp Asp	95	100	105
Leu Gly Trp Phe Val Cys Lys Ser Ser Asp Trp Phe Ile His Thr	110	115	120
Cys Met Ala Ala Lys Ser Leu Thr Ile Val Val Val Ala Lys Val	125	130	135
Cys Phe Met Tyr Ala Ser Asp Pro Ala Lys Gln Val Ser Ile His	140	145	150
Asn Tyr Thr Ile Trp Ser Val Leu Val Ala Ile Trp Thr Val Ala	155	160	165
Ser Leu Leu Pro Leu Pro Glu Trp Phe Phe Ser Thr Ile Arg His	170	175	180
His Glu Gly Val Glu Met Cys Leu Val Asp Val Pro Ala Val Ala	185	190	195
Glu Glu Phe Met Ser Met Phe Gly Lys Leu Tyr Pro Leu Leu Ala	200	205	210
Phe Gly Leu Pro Leu Phe Phe Ala Ser Phe Tyr Phe Trp Arg Ala	215	220	225
Tyr Asp Gln Cys Lys Lys Arg Gly Thr Lys Thr Gln Asn Leu Arg	230	235	240
Asn Gln Ile Arg Ser Lys Gln Val Thr Val Met Leu Leu Ser Ile	245	250	255
Ala Ile Ile Ser Ala Leu Leu Trp Leu Pro Glu Trp Val Ala Trp	260	265	270
Leu Trp Val Trp His Leu Lys Ala Ala Gly Pro Ala Pro Pro Gln			

	275		280		285
Gly Phe Ile Ala	Leu Ser Gln Val Leu	Met Phe Ser Ile Ser	Ser		
	290		295		300
Ala Asn Pro Leu	Ile Phe Leu Val Met	Ser Glu Glu Phe Arg	Glu		
	305		310		315
Gly Leu Lys Gly	Val Trp Lys Trp Met	Ile Thr Lys Lys Pro	Pro		
	320		325		330
Thr Val Ser Glu	Ser Gln Glu Thr Pro	Ala Gly Asn Ser Glu	Gly		
	335		340		345
Leu Pro Asp Lys	Val Pro Ser Pro Glu	Ser Pro Ala Ser Ile	Pro		
	350		355		360
Glu Lys Glu Lys	Pro Ser Ser Pro Ser	Ser Gly Lys Gly Lys	Thr		
	365		370		375
Glu Lys Ala Glu	Ile Pro Ile Leu Pro	Asp Val Glu Gln Phe	Trp		
	380		385		390
His Glu Arg Asp	Thr Val Pro Ser Val	Gln Asp Asn Asp Pro	Ile		
	395		400		405
Pro Trp Glu His	Glu Asp Gln Glu Thr	Gly Glu Gly Val Lys			
	410		415		

<210> 12

<211> 314

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472072CD1

<400> 12

Met Gly Asp Val	Asn Gln Ser Val Ala	Ser Asp Phe Ile Leu	Val
1	5	10	15
Gly Leu Phe Ser	His Ser Gly Ser Arg	Gln Leu Leu Phe Ser	Leu
	20	25	30
Val Ala Val Met	Phe Val Ile Gly Leu	Leu Gly Asn Thr Val	Leu
	35	40	45
Leu Phe Leu Ile	Arg Val Asp Ser Arg	Leu His Thr Pro Met	Tyr
	50	55	60
Phe Leu Leu Ser	Gln Leu Ser Leu Phe	Asp Ile Gly Cys Pro	Met
	65	70	75
Val Thr Ile Pro	Lys Met Ala Ser Asp	Phe Leu Arg Gly Glu	Gly
	80	85	90
Ala Thr Ser Tyr	Gly Gly Gly Ala Ala	Gln Ile Phe Phe Leu	Thr
	95	100	105
Leu Met Gly Val	Ala Glu Gly Val Leu	Leu Val Leu Met Ser	Tyr
	110	115	120
Asp Arg Tyr Val	Ala Val Cys Gln Pro	Leu Gln Tyr Pro Val	Leu
	125	130	135
Met Arg Arg Gln	Val Cys Leu Leu Met	Met Gly Ser Ser Trp	Val
	140	145	150
Val Gly Val Leu	Asn Ala Ser Ile Gln	Thr Ser Ile Thr Leu	His
	155	160	165
Phe Pro Tyr Cys	Ala Ser Arg Ile Val	Asp His Phe Phe Cys	Glu
	170	175	180
Val Pro Ala Leu	Leu Lys Leu Ser Cys	Ala Asp Thr Cys Ala	Tyr
	185	190	195
Glu Met Ala Leu	Ser Thr Ser Gly Val	Leu Ile Leu Met Leu	Pro
	200	205	210
Leu Ser Leu Ile	Ala Thr Ser Tyr Gly	His Val Leu Gln Ala	Val
	215	220	225
Leu Ser Met Arg	Ser Glu Glu Ala Arg	His Lys Ala Val Thr	Thr
	230	235	240
Cys Ser Ser His	Ile Thr Val Val Gly	Leu Phe Tyr Gly Ala	Ala
	245	250	255
Val Phe Met Tyr	Met Val Pro Cys Ala	Tyr His Ser Pro Gln	Gln
	260	265	270
Asp Asn Val Val	Ser Leu Phe Tyr Ser	Leu Val Thr Pro Thr	Leu

Asn Pro Leu Ile	275	Tyr Ser Leu Arg Asn	280	Pro Glu Val Trp Met	285
Leu Val Lys Val	290	Leu Ser Arg Ala Gly	295	Leu Arg Gln Met Cys	300
	305		310		

<210> 13
 <211> 254
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472074CD1

<400> 13

Met Ala Ser Arg Tyr	Val Ala Val Gly	Met Ile Leu Ser Gln Thr
1	5	10
Val Val Gly Val Leu	Gly Ser Phe Ser Val	Leu Leu His Tyr Leu
20	25	30
Ser Phe Tyr Cys Thr	Gly Cys Arg Leu Arg	Ser Thr Asp Leu Ile
35	40	45
Val Lys His Leu Ile	Val Ala Asn Phe Leu	Ala Leu Arg Cys Lys
50	55	60
Gly Val Pro Gln Thr	Met Ala Ala Phe Gly	Val Arg Tyr Phe Leu
65	70	75
Asn Ala Leu Gly Cys	Lys Leu Val Phe Tyr	Leu His Arg Val Gly
80	85	90
Arg Gly Val Ser Ile	Gly Thr Thr Cys Leu	Leu Ser Val Phe Gln
95	100	105
Val Ile Thr Val Ser	Ser Arg Lys Ser Arg	Trp Ala Lys Leu Lys
110	115	120
Glu Lys Ala Pro Lys	His Val Gly Phe Ser	Val Leu Leu Cys Trp
125	130	135
Ile Val Cys Met Leu	Val Asn Ile Ile Phe	Pro Met Tyr Val Thr
140	145	150
Gly Lys Trp Asn Tyr	Thr Asn Ile Thr Val	Asn Glu Asp Leu Gly
155	160	165
Tyr Cys Ser Gly Gly	Gly Asn Asn Lys Ile	Ala Gln Thr Leu Arg
170	175	180
Ala Met Leu Leu Ser	Phe Pro Asp Val Leu	Cys Leu Gly Leu Met
185	190	195
Leu Trp Val Ser Ser	Ser Met Val Cys Ile	Leu His Arg His Lys
200	205	210
Gln Arg Val Gln His	Ile Asp Arg Ser Asp	Leu Ser Pro Arg Ala
215	220	225
Ser Pro Glu Asn Arg	Ala Thr Gln Ser Ile	Leu Ile Leu Val Ser
230	235	240
Thr Phe Val Ser Ser	Tyr Thr Leu Ser Cys	Leu Phe Gln Val
245	250	

<210> 14
 <211> 362
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472077CD1

<400> 14

Met Tyr Lys Asp Cys	Ile Glu Ser Thr Gly	Asp Tyr Phe Leu Leu
1	5	10
Cys Asp Ala Glu Gly	Pro Trp Gly Ile Ile	Leu Glu Ser Leu Ala
20	25	30
Ile Leu Gly Ile Val	Val Thr Ile Leu Leu	Leu Leu Ala Phe Leu
35	40	45

Phe	Leu	Met	Arg	Lys	Ile	Gln	Asp	Cys	Ser	Gln	Trp	Asn	Val	Leu	
				50					55					60	
Pro	Thr	Gln	Leu	Leu	Phe	Leu	Leu	Ser	Val	Leu	Gly	Leu	Phe	Gly	
				65					70					75	
Leu	Ala	Phe	Ala	Phe	Ile	Ile	Glu	Leu	Asn	Gln	Gln	Thr	Ala	Pro	
				80					85					90	
Val	Arg	Tyr	Phe	Leu	Phe	Gly	Val	Leu	Phe	Ala	Leu	Cys	Phe	Ser	
				95					100					105	
Cys	Leu	Leu	Ala	His	Ala	Ser	Asn	Leu	Val	Lys	Leu	Val	Arg	Gly	
				110					115					120	
Cys	Val	Ser	Phe	Ser	Trp	Thr	Thr	Ile	Leu	Cys	Ile	Ala	Ile	Gly	
				125					130					135	
Cys	Ser	Leu	Leu	Gln	Ile	Ile	Ile	Ala	Thr	Glu	Tyr	Val	Thr	Leu	
				140					145					150	
Ile	Met	Thr	Arg	Gly	Met	Met	Phe	Val	Asn	Met	Thr	Pro	Cys	Gln	
				155					160					165	
Leu	Asn	Val	Asp	Phe	Val	Val	Leu	Leu	Val	Tyr	Val	Leu	Phe	Leu	
				170					175					180	
Met	Ala	Leu	Thr	Phe	Phe	Val	Ser	Lys	Ala	Thr	Phe	Cys	Gly	Pro	
				185					190					195	
Cys	Glu	Asn	Trp	Lys	Gln	His	Gly	Arg	Leu	Ile	Phe	Ile	Thr	Val	
				200					205					210	
Leu	Phe	Ser	Ile	Ile	Ile	Trp	Val	Val	Trp	Ile	Ser	Met	Leu	Leu	
				215					220					225	
Arg	Gly	Asn	Pro	Gln	Phe	Gln	Arg	Gln	Pro	Gln	Trp	Asp	Asp	Pro	
				230					235					240	
Val	Val	Cys	Ile	Ala	Leu	Val	Thr	Asn	Ala	Trp	Val	Phe	Leu	Leu	
				245					250					255	
Leu	Tyr	Ile	Val	Pro	Glu	Leu	Cys	Ile	Leu	Tyr	Arg	Ser	Cys	Arg	
				260					265					270	
Gln	Glu	Cys	Pro	Leu	Gln	Gly	Asn	Ala	Cys	Pro	Val	Thr	Ala	Tyr	
				275					280					285	
Gln	His	Ser	Phe	Gln	Val	Glu	Asn	Gln	Glu	Leu	Ser	Arg	Asp	Lys	
				290					295					300	
Trp	Lys	Val	Leu	Leu	Asn	Ser	Asp	Phe	Leu	Ser	His	Ser	Gly	Ala	
				305					310					315	
Ala	Arg	Asp	Ser	Asp	Gly	Ala	Glu	Glu	Asp	Val	Ala	Leu	Thr	Ser	
				320					325					330	
Tyr	Gly	Thr	Pro	Ile	Gln	Pro	Gln	Thr	Val	Asp	Pro	Thr	Gln	Glu	
				335					340					345	
Cys	Phe	Ile	Pro	Gln	Ala	Lys	Leu	Ser	Pro	Gln	Gln	Asp	Ala	Gly	
				350					355					360	
Gly	Val														

<210> 15

<211> 370

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472082CD1

<400> 15

Met	Cys	Lys	Cys	Phe	Arg	Ser	Gly	Asn	Ser	Thr	Pro	Val	Leu	Cys	
				5					10					15	
His	Arg	Asn	Ser	Glu	Ala	Trp	Gln	Pro	Arg	Lys	Ala	Pro	Arg	Thr	
				20					25					30	
Gln	Gln	Thr	Asp	Met	Gly	Tyr	Thr	Asn	Leu	Asn	Ser	Lys	Lys	Glu	
				35					40					45	
Cys	Met	Tyr	Ile	Lys	Glu	Asn	Phe	Lys	Lys	Thr	Val	Asp	Lys	Ile	
				50					55					60	
Val	Asp	Pro	Gly	Asn	His	Ser	Ser	Val	Thr	Glu	Ser	Ile	Leu	Ala	
				65					70					75	
Gly	Leu	Ser	Glu	Gln	Pro	Glu	Leu	Gln	Leu	Arg	Leu	Phe	Leu	Leu	
				80					85					90	

```

Phe Leu Gly Ile Cys Val Val Thr Val Val Gly Asn Leu Gly Met
  95 100 105
Ile Thr Leu Ile Gly Leu Ser Ser His Leu His Thr Pro Met Tyr
 110 115 120
Tyr Phe Leu Ser Ser Leu Ser Phe Ile Asp Phe Cys His Ser Thr
 125 130 135
Val Ile Thr Pro Lys Met Leu Val Asn Phe Ala Thr Glu Lys Asn
 140 145 150
Ile Ile Ser Tyr Pro Glu Cys Met Ala Gln Leu Tyr Leu Phe Ser
 155 160 165
Ile Phe Ala Ile Ala Glu Cys His Met Leu Ala Ala Met Ala Tyr
 170 175 180
Asp Cys Tyr Val Ala Ile Cys Ser Pro Leu Leu Tyr Asn Val Ile
 185 190 195
Met Ser Tyr His His Cys Phe Trp Leu Thr Val Gly Val Tyr Ile
 200 205 210
Leu Gly Ile Leu Gly Ser Thr Ile His Thr Ser Phe Met Leu Arg
 215 220 225
Leu Phe Leu Cys Lys Thr Asn Val Ile Asn His Tyr Phe Cys Asp
 230 235 240
Leu Phe Pro Leu Leu Gly Leu Ser Cys Ser Ser Thr Tyr Ile Asn
 245 250 255
Glu Leu Leu Val Leu Val Leu Ser Ala Phe Asn Ile Leu Met Pro
 260 265 270
Ala Leu Thr Ile Leu Ala Ser Tyr Ile Phe Ile Ile Ala Ser Ile
 275 280 285
Leu Arg Ile His Ser Thr Glu Gly Arg Ser Lys Ala Phe Ser Thr
 290 295 300
Cys Ser Ser His Ile Leu Ala Val Ala Val Phe Phe Gly Ser Ala
 305 310 315
Ala Phe Met Tyr Leu Gln Pro Ser Ser Val Ser Ser Met Asp Gln
 320 325 330
Arg Lys Val Ser Ser Val Phe Tyr Thr Thr Ile Val Pro Met Leu
 335 340 345
Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys Leu Ala
 350 355 360
Val Lys Lys Ile Leu His Gln Thr Ala Cys
 365 370

```

<210> 16
 <211> 319
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472128CD1

```

<400> 16
Met Thr Pro Gly Glu Leu Ala Leu Ala Ser Gly Asn His Thr Pro
  1  5 10 15
Val Thr Lys Phe Ile Leu Gln Gly Phe Ser Asn Tyr Pro Asp Leu
 20 25 30
Gln Glu Leu Leu Phe Gly Ala Ile Leu Leu Ile Tyr Ala Ile Thr
 35 40 45
Val Val Gly Asn Leu Gly Met Met Ala Leu Ile Phe Thr Asp Ser
 50 55 60
His Leu Gln Ser Pro Met Tyr Phe Phe Leu Asn Val Leu Ser Phe
 65 70 75
Leu Asp Ile Cys Tyr Ser Ser Val Val Thr Pro Lys Leu Leu Val
 80 85 90
Asn Phe Leu Val Ser Asp Lys Ser Ile Ser Phe Glu Gly Cys Val
 95 100 105
Val Gln Leu Ala Phe Phe Val Val His Val Thr Ala Glu Ser Phe
 110 115 120
Leu Leu Ala Ser Met Ala Tyr Asp Arg Phe Leu Ala Ile Cys Gln
 125 130 135

```

```

Pro Leu His Tyr Gly Ser Ile Met Thr Arg Gly Thr Cys Leu Gln
140 145 150
Leu Val Ala Val Ser Tyr Ala Phe Gly Gly Ala Asn Ser Ala Ile
155 160 165
Gln Thr Gly Asn Val Phe Ala Leu Pro Phe Cys Gly Pro Asn Gln
170 175 180
Leu Thr His Tyr Tyr Cys Asp Ile Pro Pro Leu Leu His Leu Ala
185 190 195
Cys Ala Asn Thr Ala Thr Ala Arg Val Val Leu Tyr Val Phe Ser
200 205 210
Ala Leu Val Thr Leu Leu Pro Ala Ala Val Ile Leu Thr Ser Tyr
215 220 225
Cys Leu Val Leu Val Ala Ile Gly Arg Met Arg Ser Val Ala Gly
230 235 240
Arg Glu Lys Asp Leu Ser Thr Cys Ala Ser His Phe Leu Ala Ile
245 250 255
Ala Ile Phe Tyr Gly Thr Val Val Phe Thr Tyr Val Gln Pro His
260 265 270
Gly Ser Thr Asn Asn Thr Asn Gly Gln Val Val Ser Val Phe Tyr
275 280 285
Thr Ile Ile Ile Pro Met Leu Asn Pro Phe Ile Tyr Ser Leu Arg
290 295 300
Asn Lys Glu Val Lys Gly Ala Leu Gln Arg Lys Leu Gln Val Asn
305 310 315
Ile Phe Pro Gly

```

```

<210> 17
<211> 312
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472134CD1

```

```

<400> 17
Met Asp Thr Gly Asn Trp Ser Gln Val Ala Glu Phe Ile Ile Leu
1 5 10 15
Gly Phe Pro His Leu Gln Gly Val Gln Ile Tyr Leu Phe Leu Leu
20 25 30
Leu Leu Leu Ile Tyr Leu Met Thr Val Leu Gly Asn Leu Leu Ile
35 40 45
Phe Leu Val Val Cys Leu Asp Ser Arg Leu His Thr Pro Met Tyr
50 55 60
His Phe Val Ser Ile Leu Ser Phe Ser Glu Leu Gly Tyr Thr Ala
65 70 75
Ala Thr Ile Pro Lys Met Leu Ala Asn Leu Leu Ser Glu Lys Lys
80 85 90
Thr Ile Ser Phe Ser Gly Cys Leu Leu Gln Ile Tyr Phe Phe His
95 100 105
Ser Leu Gly Ala Thr Glu Cys Tyr Leu Leu Thr Ala Met Ala Tyr
110 115 120
Asp Arg Tyr Leu Ala Ile Cys Arg Pro Leu His Tyr Pro Thr Leu
125 130 135
Met Thr Pro Thr Leu Cys Ala Glu Ile Ala Ile Gly Cys Trp Leu
140 145 150
Gly Gly Leu Ala Gly Pro Val Val Glu Ile Ser Leu Ile Ser Arg
155 160 165
Leu Pro Phe Cys Gly Pro Asn Arg Ile Gln His Val Phe Cys Asp
170 175 180
Phe Pro Pro Val Leu Ser Leu Ala Cys Thr Asp Thr Ser Ile Asn
185 190 195
Val Leu Val Asp Phe Val Ile Asn Ser Cys Lys Ile Leu Ala Thr
200 205 210
Phe Leu Leu Ile Leu Cys Ser Tyr Val Gln Ile Ile Cys Thr Val
215 220 225

```


Leu	Arg	Ile	Pro	Ser	Ala	Ala	Gly	Lys	Arg	Lys	Ala	Ile	Ser	Thr		
				230					235					240		
Cys	Ala	Ser	His	Phe	Thr	Val	Val	Leu	Ile	Phe	Tyr	Gly	Ser	Ile		
				245					250					255		
Leu	Ser	Met	Tyr	Val	Gln	Leu	Lys	Lys	Ser	Tyr	Ser	Leu	Asp	Tyr		
				260					265					270		
Asp	Gln	Ala	Leu	Ala	Val	Val	Tyr	Ser	Val	Leu	Thr	Pro	Phe	Leu		
				275					280					285		
Asn	Pro	Phe	Ile	Tyr	Ser	Leu	Arg	Asn	Lys	Glu	Ile	Lys	Glu	Ala		
				290					295					300		
Val	Arg	Arg	Gln	Leu	Lys	Arg	Ile	Gly	Ile	Leu	Ala					
				305					310							

<210> 18
 <211> 321
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472136CD1

<400> 18

Met	Asn	Gln	Thr	Leu	Asn	Ser	Ser	Gly	Thr	Val	Glu	Ser	Ala	Leu		
1				5					10					15		
Asn	Tyr	Ser	Arg	Gly	Ser	Thr	Val	His	Thr	Ala	Tyr	Leu	Val	Leu		
				20					25					30		
Ser	Ser	Leu	Ala	Met	Phe	Thr	Cys	Leu	Cys	Gly	Met	Ala	Gly	Asn		
				35					40					45		
Ser	Met	Val	Ile	Trp	Leu	Leu	Gly	Phe	Arg	Met	His	Arg	Asn	Pro		
				50					55					60		
Phe	Cys	Ile	Tyr	Ile	Leu	Asn	Leu	Ala	Ala	Ala	Asp	Leu	Leu	Phe		
				65					70					75		
Leu	Phe	Ser	Met	Ala	Ser	Thr	Leu	Ser	Leu	Glu	Thr	Gln	Pro	Leu		
				80					85					90		
Val	Asn	Thr	Thr	Asp	Lys	Val	His	Glu	Leu	Met	Lys	Arg	Leu	Met		
				95					100					105		
Tyr	Phe	Ala	Tyr	Thr	Val	Gly	Leu	Ser	Leu	Leu	Thr	Ala	Ile	Ser		
				110					115					120		
Thr	Gln	Arg	Cys	Leu	Ser	Val	Leu	Phe	Pro	Ile	Trp	Phe	Lys	Cys		
				125					130					135		
His	Arg	Pro	Arg	His	Leu	Ser	Ala	Trp	Val	Cys	Gly	Leu	Leu	Trp		
				140					145					150		
Thr	Leu	Cys	Leu	Leu	Met	Asn	Gly	Leu	Thr	Ser	Ser	Phe	Cys	Ser		
				155					160					165		
Lys	Phe	Leu	Lys	Phe	Asn	Glu	Asp	Arg	Cys	Phe	Arg	Val	Asp	Met		
				170					175					180		
Val	Gln	Ala	Ala	Leu	Ile	Met	Gly	Val	Leu	Thr	Pro	Val	Met	Thr		
				185					190					195		
Leu	Ser	Ser	Leu	Thr	Leu	Phe	Val	Trp	Val	Arg	Arg	Ser	Ser	Gln		
				200					205					210		
Gln	Trp	Arg	Arg	Gln	Pro	Thr	Arg	Leu	Phe	Val	Val	Val	Leu	Ala		
				215					220					225		
Ser	Val	Leu	Val	Phe	Leu	Ile	Cys	Ser	Leu	Pro	Leu	Ser	Ile	Tyr		
				230					235					240		
Trp	Phe	Val	Leu	Tyr	Trp	Leu	Ser	Leu	Pro	Glu	Met	Gln	Val			
				245					250					255		
Leu	Cys	Phe	Ser	Leu	Ser	Arg	Leu	Ser	Ser	Ser	Val	Ser	Ser	Ser		
				260					265					270		
Ala	Asn	Pro	Val	Ile	Tyr	Phe	Leu	Val	Gly	Ser	Arg	Arg	Ser	His		
				275					280					285		
Arg	Leu	Pro	Thr	Arg	Ser	Leu	Gly	Thr	Val	Leu	Gln	Gln	Ala	Leu		
				290					295					300		
Arg	Glu	Glu	Pro	Glu	Leu	Glu	Gly	Gly	Glu	Thr	Pro	Thr	Val	Gly		
				305					310					315		
Thr	Asn	Glu	Met	Gly	Ala											
				320												

<210> 19
 <211> 316
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472142CD1

<400> 19
 Met Gln Gly Glu Asn Phe Thr Ile Trp Ser Ile Phe Phe Leu Glu
 1 5 10 15
 Gly Phe Ser Gln Tyr Pro Gly Leu Glu Val Val Leu Phe Val Phe
 20 25 30
 Ser Leu Val Met Tyr Leu Thr Thr Leu Leu Gly Asn Ser Thr Leu
 35 40 45
 Ile Leu Ile Thr Ile Leu Asp Ser Arg Leu Lys Thr Pro Met Tyr
 50 55 60
 Leu Phe Leu Gly Asn Leu Ser Phe Met Asp Ile Cys Tyr Thr Ser
 65 70 75
 Ala Ser Val Pro Thr Leu Leu Val Asn Leu Leu Ser Ser Gln Lys
 80 85 90
 Thr Ile Ile Phe Ser Gly Cys Ala Val Gln Met Tyr Leu Ser Leu
 95 100 105
 Ala Met Gly Ser Thr Glu Cys Val Leu Leu Ala Val Met Ala Tyr
 110 115 120
 Asp Arg Tyr Val Ala Ile Cys Asn Pro Leu Arg Tyr Ser Ile Ile
 125 130 135
 Met Asn Arg Cys Val Cys Ala Arg Met Ala Thr Val Ser Trp Val
 140 145 150
 Thr Gly Cys Leu Thr Ala Leu Leu Glu Thr Ser Phe Ala Leu Gln
 155 160 165
 Ile Pro Leu Cys Gly Asn Leu Ile Asp His Phe Thr Cys Glu Ile
 170 175 180
 Leu Ala Val Leu Lys Leu Ala Cys Thr Ser Ser Leu Leu Met Asn
 185 190 195
 Thr Ile Met Leu Val Val Ser Ile Leu Leu Leu Pro Ile Pro Met
 200 205 210
 Leu Leu Val Cys Ile Ser Tyr Ile Phe Ile Leu Ser Thr Ile Leu
 215 220 225
 Arg Ile Thr Ser Ala Glu Gly Arg Asn Lys Ala Phe Ser Thr Cys
 230 235 240
 Gly Ala His Leu Thr Val Val Ile Leu Tyr Tyr Gly Ala Ala Leu
 245 250 255
 Ser Met Tyr Leu Lys Pro Ser Ser Ser Asn Ala Gln Lys Ile Asp
 260 265 270
 Lys Ile Ile Ser Leu Leu Tyr Gly Val Leu Thr Pro Met Leu Asn
 275 280 285
 Pro Ile Ile Tyr Ser Leu Arg Asn Lys Glu Val Lys Asp Ala Met
 290 295 300
 Lys Lys Leu Leu Gly Lys Ile Thr Leu His Gln Thr His Glu His
 305 310 315
 Leu

<210> 20
 <211> 325
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472171CD1

<400> 20
 Met Glu Pro Leu Asn Arg Thr Glu Val Ser Glu Phe Phe Leu Lys
 1 5 10 15

<400>	21													
Met	Gly	Asn	Trp	Ser	Thr	Val	Thr	Glu	Ile	Thr	Leu	Ile	Ala	Phe
1				5					10					15
Pro	Ala	Leu	Leu	Glu	Ile	Arg	Ile	Ser	Leu	Phe	Val	Val	Leu	Val
				20					25					30
Val	Thr	Tyr	Thr	Leu	Thr	Ala	Thr	Gly	Asn	Ile	Thr	Ile	Ile	Ser
				35					40					45
Leu	Ile	Trp	Ile	Asp	His	Arg	Leu	Gln	Thr	Pro	Met	Tyr	Phe	Phe
				50					55					60
Leu	Ser	Asn	Leu	Ser	Phe	Leu	Asp	Ile	Leu	Tyr	Thr	Thr	Val	Ile
				65					70					75
Thr	Pro	Lys	Leu	Leu	Ala	Cys	Leu	Leu	Gly	Glu	Glu	Lys	Thr	Ile
				80					85					90
Ser	Phe	Ala	Gly	Cys	Met	Ile	Gln	Thr	Tyr	Phe	Tyr	Phe	Phe	Leu
				95					100					105

Gly	Thr	Val	Glu	Phe	Ile	Leu	Leu	Ala	Val	Met	Ser	Phe	Asp	Arg	
				110					115					120	
Tyr	Met	Ala	Ile	Cys	Asp	Pro	Leu	His	Tyr	Thr	Val	Ile	Met	Asn	
				125					130					135	
Ser	Arg	Ala	Cys	Leu	Leu	Leu	Val	Leu	Gly	Cys	Trp	Val	Gly	Ala	
				140					145					150	
Phe	Leu	Ser	Val	Leu	Phe	Pro	Thr	Ile	Val	Val	Thr	Arg	Leu	Pro	
				155					160					165	
Tyr	Cys	Arg	Lys	Glu	Ile	Asn	His	Phe	Phe	Cys	Asp	Ile	Ala	Pro	
				170					175					180	
Leu	Leu	Gln	Val	Ala	Cys	Ile	Asn	Thr	His	Leu	Ile	Glu	Lys	Ile	
				185					190					195	
Asn	Phe	Leu	Leu	Ser	Ala	Leu	Val	Ile	Leu	Ser	Ser	Leu	Ala	Phe	
				200					205					210	
Thr	Thr	Gly	Ser	Tyr	Val	Tyr	Ile	Ile	Ser	Thr	Ile	Leu	Arg	Ile	
				215					220					225	
Pro	Ser	Thr	Gln	Gly	Arg	Gln	Lys	Ala	Phe	Ser	Thr	Cys	Ala	Ser	
				230					235					240	
His	Ile	Thr	Val	Val	Ser	Ile	Ala	His	Gly	Ser	Asn	Ile	Phe	Val	
				245					250					255	
Tyr	Val	Arg	Pro	Asn	Gln	Asn	Ser	Ser	Leu	Asp	Tyr	Asp	Lys	Val	
				260					265					270	
Ala	Ala	Val	Leu	Ile	Thr	Val	Val	Thr	Pro	Leu	Leu	Asn	Pro	Phe	
				275					280					285	
Ile	Tyr	Ser	Leu	Arg	Asn	Glu	Lys	Val	Gln	Glu	Val	Leu	Arg	Glu	
				290					295					300	
Thr	Val	Asn	Arg	Ile	Met	Thr	Leu	Ile	Gln	Arg	Lys	Thr			
				305					310						

<210> 22

<211> 1413

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472033CB1

<220>

<221> unsure

<222> 176, 383, 927

<223> a, t, c, g, or other

<400> 22

atgaatcaga	cggagcccg	ccagctggca	gatgggggagc	atctgagtgg	atacgcacg	60
agcagcaaca	gcgtgcgcta	tctggacgac	cgccatccgc	tgactacct	tgacctgggc	120
acggtgcacg	ccctcaacac	cactgccatc	aacacctcgg	atctgaatga	gactgngagc	180
aggccgctgg	acccggtgct	tatcgatagg	ttcctgagca	acagggcggg	ggacagcccc	240
tggtaccaca	tgctcatcag	catgtacggc	gtgctaatac	tcttcggcgc	cctaggcaac	300
acccttggtt	gttatagccc	gtcatccgga	agcccatcat	gcgcactgct	cgcaatctgg	360
ttcatcctca	acctggccat	atnccggcaa	agcaagtgtg	agtctcatcc	gagcggactt	420
tcagacctac	ttttatgcct	agtcaccatg	ccgctgacct	tgatggagat	cctgtccaag	480
tactggccct	acggctcctg	ctccatcctg	tgcaaaacga	ttgccatgct	gcaggcactt	540
tgtattttcg	tgtcgacaat	atccataacg	gccattgcct	tcgacagata	tcagggtgatc	600
gtgtacccca	cgccgggacag	cctgcagttc	gtggggcgcg	tgacgatcct	ggcgggggatc	660
tgggcactgg	cactgctgct	ggcctcgccg	ctgttcgtct	acaaggagct	gatcaacaca	720
gacacgcggg	cactcctgca	gcagatcggc	ctgcaggaca	cgatcccgta	ctgcattgag	780
gactggccaa	gtcgcaacgg	gcgcttctac	tactcgatct	tctcgctgtg	cgtacaatac	840
ctggtgcccc	tcctgatcgt	ctcggtgcca	tacttcggga	tctacaacaa	gctgaagagc	900
cgcatcacgg	tggtggctgt	gcaggcncgg	aagggtggagc	gggggcgggc	gatgaagcgc	960
accaactgcc	tactgatcag	catcgccatc	atctttggcg	tttcttggtt	gccgcttgac	1020
tttttcaacc	tgtacgcgga	catggagcgc	tcgcccgtca	ctcagagcat	gctagtccgc	1080
tacgccatct	gccacatgat	cggcatgagc	tccgcctgct	ccaaccggtt	gctctacggc	1140
tggtctcaacg	acaacttccg	taaagaaatt	caagaactgc	tctgccgttg	ctcagacact	1200
aatgtggctc	ttaacggtea	cacgacaggc	tgcaacgtcc	aggcggcggc	gcgcaggcgt	1260
cgcaagtatg	ggcgccgaat	tctccaaagg	cgaactcaag	ctgctggggc	aggcggcgcc	1320
agagcgggtac	cgccggcgggg	gaggcggtct	ggcggccacc	gacttcatga	ccggcaccac	1380

gaggggggac tcgccaacat agttcatcat tga

1413

<210> 23
<211> 1076
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472034CB1

<400> 23
tctacccaaa tgagaaggaa cagagaaaca attctggtaa tattacaaaa caaggttctt 60
taacacctcc gaaggatcaa accagctcac cagcaatgga tccaaaccaa gatgaaatct 120
ctgaattacc agaaaaagaa ttcagaagat caattattaa gctgatcaaa gaggcaccag 180
aaaaagggat tccaggttta gaggaagcc agcactggat tgcactgcc ctgggcatcc 240
tttacctcct tgcttttagtg ggcaatgtta ccattctctt catcatctgg atggacccat 300
ccttgcacca atctatgtac ctcttcctgt ccatgctagc tgccatcgac ctggttcttg 360
cctctccac tgcacccaaa gcccttgcat tgctcctggt tcatgcccac gagattgggt 420
acatcgctcg cctgatccag atgttcttca tccatgcatt ctctccatg gagtccagg 480
tacttggtgc catggctctg gatcgctatg tagccatttg tcccccttg caccattcca 540
caatctgca tccaggggtc atagggcgca tcggaatggt ggtgctggtg aggggattac 600
tactccttat ccccttcccc attttgttgg gaacacttat cttctgcaa gccaccatca 660
taggcgatgc ctattgtgaa catatggctg ttgtgaaact tgcctgctca gaaaccacag 720
tcaatcgagc ttatgggctg actatggcct tgcttgatg tgggctggat gttctggcca 780
ttggtgtttc ctatgccac atcctccagg cagtgcgtgaa ggtaccagg agtgaggccc 840
gacttaaggc gtttagcaca tgtggctctc atatttgtgt catcctggtc ttctatgtcc 900
ctggaatttt ctcttctc actcaccgct ttggtcatca tgtacccat catgtccatg 960
ttcttctggc cacacgggat ctctcatgc cactgcgct caatcctctt gtctatggag 1020
tgaagactca gcagatccgc cagcgagtgc tcagagtgtt tacacaaaag gattga 1076

<210> 24
<211> 948
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472035CB1

<400> 24
atggaaaccc ctgcctcctt cctccttggt ggtatcccag gactgcaatc ttcacatctt 60
tggtctggcta tctcactgag tgccatgtac atcatagccc tgtaggaaa caccatcatc 120
gtgactgcaa tctggatgga ttccactcgg catgagccca tgtattgctt tctgtgtgtt 180
ctggctgctg tggacattgt tatggcctcc tcgggtgtac ccaagatggt gagcatcttc 240
tgctcaggag acagctcaat cagctttagt gcttggttca ctccagatgtt tttgtccac 300
ttagccacag ctgtggagac ggggctgctg ctgaccatgg cttttgaccg ctatgtagcc 360
atctgcaagc ctctacacta caagagaatt ctacgcctc aagtgtgctt gggatgagt 420
atggccatca ccatcagagc tatcatagcc ataactccac tgagttggat ggtgagtcac 480
ctacctttct gtggctccaa tgtggttgct cactcctact gtgagcacat agctttggcc 540
aggtagcat gtgctgaccc cgtgccacgc agtctctaca gtctgattgg ttctctctt 600
atggtgggct ctgatgtggc cttcattgct gcctctata tcttaattct caaggcagta 660
tttggctctt cctcaaagac tgctcagttg aaagcattaa gcacatgtgg ctcccatgtg 720
ggggttatgg ctttgtacta tctacctggg atggcatcca tctatcgagg ctggttgggg 780
caggatgtag tgcccttgca caccacagtc ctgctagctg acctgtacgt gatcatccca 840
gccaccttaa atcccatcat ctatggcatg aggaccaaac aactgcggga gagaatatgg 900
agttatctga tgcattgtcct ctttgaccat tccaacctgg gttcatga 948

<210> 25
<211> 945
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472036CB1

```

<400> 25
atgtcagcct ccaatatcac cttaacacat ccaactgcct tcttgttggt ggggattcca 60
ggcctggaac acctgcacat ctggatctcc atccctttct gcttagcata tacactggcc 120
ctgcttgga actgcactct ccttctcatc atccaggctg atgcagccct ccatgaacc 180
atgtacctct ttctggccat gttggcagcc atcgacctgg tcttttctc ctcagcactg 240
cccaaatgc ttgccatatt ctggttcagg gatcgaggaga taaacttctt tgcctgtctg 300
gccagatgt tcttcttca ctccttctcc atcatggagt cagcagtgcg gctggccatg 360
gcctttgacc gctatgtggc tatctgcaag ccactgcact acaccaaggc cctgactggg 420
tccctcatca ccaagattgg catggctgct gtggcccggt ctgtgacact aatgactcca 480
ctcccttcc tgtcgagatg tttccactac tgccgaggcc cagtgatcgc tactgtctac 540
tgtgaacaca tggctgtggt gaggctggcg tgtggggaca ctagcttcaa caatatctat 600
ggcatcgctg tggccatgtt tattgtggtg ttggacctgc tcttgttat cctgtcttat 660
atctttattc ttcaggcagt tctactgctt gcctctcagg aggcccgcta caaggcattt 720
gggacatgtg tctctcatat aggtgccatc ttgaccttct acacaactgt ggtcatctct 780
tcagtcatgc accgtgtagc ccgcatgct gccctcatg tccacatcct ccttgccaat 840
ttctatctgc tcttcccacc catggtcaat ccataatct atggtgtcaa gaccaagcaa 900
atccgtgaga gcatcttggg agtattccca agaaaggata tgtag 945

```

```

<210> 26
<211> 966
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472037CB1

```

```

<400> 26
atggctcatc aggcgcctga gaagcagcag gacaatggga cctggctggt gacagagttc 60
ctgctggtgg gattctccaa cctcccagaa ctgaggccca ctctcttcat cttgttctc 120
ctcacctacc tggtcacact cagtggcaat gccaccatca tcaccatcat ccagggtgat 180
cgcactctcc acacacctat gtaccgcttc ctggccgtgc tctccctctc tgagacctgc 240
tacacactgg tcaccatccc caatatgctg gctcatctgc tgatggagag ccaggccatc 300
tccatcgcggt gctgtcggtt ccagatgttc ttcttcttag gcttgggttg cagccattgt 360
ttctctctta cctgatggg ctatgacagg tatgtggcca tctgccatcc cttgcgtctac 420
tctgtgatca tgagaccac cgtctgcctg tgtttgggag ccttggtttt ctgctctggt 480
ttctcagtgg ccttgattga gacctgcatg atcttctct caccctcttg tggcgaggc 540
catgtggagg acttcttctg tgacattgct cctgtgctga agctcagctg tgatgagagc 600
tactcaagg gacttggcat cttcttcttg agcatctctg tgggtgctggt ctcttctctc 660
ttcattctcc tctctacgc cttcattgtg gctgccattg tgaggatccc ttcggcctct 720
ggcgcgcgca aagccttctc tacctgcgca gccacactca cgggtggtcat cgtacatttt 780
ggttgtgctt ccatcatcta cctgaggcgg gactctgggg ctaatccctc ccaggaccgc 840
ctggtggcgg tgtctacac cgtggtgaca ccgctgctga accctgtggt ttacaccctg 900
aggaacaagg aggtgagggg agcgtgtagg aaaaacctgg cacggggctg tggagcattt 960
aagtaa 966

```

```

<210> 27
<211> 996
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472039CB1

```

```

<400> 27
atgagtcctg atgggaacca cagtagtgat ccaacagagt tcgtcctggc agggctccca 60
aatctcaaca gcgcaagagt ggaattattt tctgtgttct ttcttgtcta tctctgaat 120
ctgacaggca atgtgttgat tgtgggggtg gtaagggctg atactcgact acagaccct 180
atgtacttct ttctgggtaa cctgtcctgc cttagagatac tgctcacttc tgtcatcatt 240
ccaaagatgc tgagcaattt cctctcaagg caacacacta ttctcttgc tgcattgtat 300
accgaattct atttctactt ctttctcggg gcctccgagt tcttactggt ggctgtcatg 360
tctgcggtac gctacctggc catctgtcat cctctgcgt acccttgct catgagtggg 420
gctgtgtgct tctgtgtggc cttggcctgc tgggtggggg gactcgtccc tgtgcttggg 480
cccagtggtg cttgtggcctt gcttcttctc tgtaagcagg gtgctgtggt acagcacttc 540
ttctgcgaca gtggccact gctccgcttg gcttgacca acaccaagaa gctggaggag 600
actgactttg tcttggcctc cctcgtcatt gtatcttctc tgctgatcac tgctgtgtcc 660

```

```

tacggcctca ttgtgctggc agtcctgagc atccccctctg cttcaggccg tcagaaggcc 720
ttctctacct gtacctccca cttgatagtg gtgacctctt tctatggaag tgccattttt 780
ctctatgtgc ggccatcgca gagtgggtct gtggacacta actgggcagt gacagtaata 840
acgacatttg tgacaccact gttgaatcca ttcatctatg ccttacgtaa tgagcaagtc 900
aagggaagctt tgaaggacat gtttaggaag gtagtggcag gcgttttagg gaatctttta 960
cttgataaat gtctcagtga gaaagcagta aagtaa 996

```

```

<210> 28
<211> 1014
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472040CB1

```

```

<400> 28
atggggaacg attctgtcag ctacgagtat ggggattaca ggcacctctc ggaccgcect 60
gtggactgcc tggatggcgc ctgcctggcc atcgaccgcg tgcgcgtggc cccgctccca 120
ctgtatgccg ccattcttct ggtgggggtg ccgggcaatg ccatggtggc ctgggtggct 180
gggaagggtg ccgcgcgag ggtgggtgcc acctggttgc tccacctggc cgtggcggat 240
ttgtctgtct gtttgtctct gccatccctg gcagtgccca ttgcccgtgg aggccactgg 300
ccgtatggtg cagtgggctg tcgggcgctg cctccatca tctgctgac catgtatgcc 360
agcgtcctgc tcctggcagc tctcagtgcc gacctctgct tcctggctct cgggectgcc 420
tgggtggtct cggttcagcg ggctgcggg gtgcagggtg cctgtggggc agcctggaca 480
ctggccttgc tgctcaccgt gccctccgcc atctaccgcc ggctgcacca ggagcacttc 540
ccagcccggc tgcagtgtgt ggtggactac ggcggtcctt ccagcaccga gaatgcggtg 600
actgccatcc ggtttctttt tggttctctg gggcccttgg tggccttggc cagctgccac 660
agtgccttcc tgtgctgggc agcccgacgc tgcggccgcg tgggcacagc cattgtggtg 720
gggttttttg tctgctgggc accctaccac ctgctggggc tgggtgtcac tgtggcggcc 780
ccgaactccg cactcctggc cagggccctg cgggctgaac cctcatcgt gggccttgcc 840
ctcgctcaca gctgcctcaa tcccatgctc ttctgtatt ttgggagggc tcaactccgc 900
cggtcactgc cagctgcctg tcaactgggc ctgaggaggt cccagggcca ggacgaaagt 960
gtggacagca agaaatccac cagccatgac ctggtctcgg agatggaggt gtag 1014

```

```

<210> 29
<211> 5122
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 4250893CB1

```

```

<400> 29
gagccagcag cccggggctc cactctgggt tctgaaagcc cattccctgc tctcgggctc 60
ctcccacccc acctcttctc agccttgcag ctcaagggtt gatctcagga gtccaggacc 120
caggagaggg aagaatctga ggaacacaga acagtgagcg ttgccacac cccatctccc 180
gtcaccacat ctcccctctg taacaccctc cctgcctggc cctggacccc atcccaggac 240
ctccctatca gctgacttct tccagtgtct tgcaggcccc tctgggctcc tccctccctt 300
ggcttttctt accactcccc ctctatcggc gtctatctgt aggtgccctg ggatttataa 360
aactgggttc cgaatgctga ataagagacg gtaagagcca aggcaaagga cagcactggt 420
ctctgcctgc ctgataccct caccacctgg gaacatcccc cagacacctt cttactccg 480
ggacagagat ggctggcgga gcctggggcc gcctggcctg ttacttggag ttctgaaga 540
aggaggagct gaaggagtgc cagcttctgc tcgccaataa agcgactccc aggagctctt 600
cgggtgagac acccgctcag ccagagaaga cgagtggcat ggaggtggcc tegtacctgg 660
tggtcagta tggggagcag cgggcctggg acctagccct ccatacctgg gacagatgg 720
ggctgaggtc actgtgcgcc caagcccagg aaggggcagg ccactctccc tcattccctt 780
acagcccaag tgaaccccac ctggggtctc ccagccaacc cactccacc gcagtgtctaa 840
tgccctggat ccatgaattg ccggcggggg gaccccaggg ctgagagaga agggttttga 900
gacagctgcc tgacacatct ggagcgcgct ctccaagcca ggagtcaccc aacgccccca 1020
aagctcttcc aagctcccca gaccatgagt ctccaagcca ggagtcaccc aacgccccca 1080
catccacagc agtgctgggg agctggggat cccacacctc gccagccta gacccagag 1140
agcaggaggg tcctgggacc caatggcctc tggatgaaac gtcaggattt tactacacag 1200
aaatcagaga aagagagaga gagaatcag agaaaggcag gccccctgg gcagcggtgg 1260
taggaacgcc cccacaggcg cacaccagcc tacagcccca ccaccacca tgggagcctt 1320
ctgtgagaga gaggcctctg tccacatggc cctggaaaaa tgaggatttt aacaaaaaat

```

tcacacagct	gctacttcta	caaagaccte	accccagaag	ccaagatccc	ctggtcaaga	1380
gaagctggcc	tgattatgtg	gaggagaatc	gaggacattt	aattgagatc	agagacttat	1440
ttggcccagg	cctggatacc	caagaaccte	gcatagtcac	actgcagggg	gctgctggaa	1500
ttgggaagtc	aacactggcc	aggcaggtga	aggaagcctg	ggggagaggc	cagctgtatg	1560
gggaccgctt	ccagcatgtc	ttctacttca	gctgcagaga	gctggcccag	tccaaggtgg	1620
tgagtctcgc	tgagctcatc	ggaaaagatg	ggacagccac	tccggctccc	attagacaga	1680
tctgtcttag	gccagagcgg	ctgctcttca	tcctcgatgg	tgtagatgag	ccaggatggg	1740
tcttgaggga	gccgagttct	gagctctgtc	tgacttgagg	ccagccacag	ccggcggatg	1800
cactgcggg	cagtttgctg	gggaaaacta	tacttcccga	ggcatccttc	ctgatcacgg	1860
ctcggaaccac	agctctgcag	aacctcattc	cttcttttga	gcaggcacgt	tgggtagagg	1920
tcttgggggt	ctctgagtc	agcaggaagg	aatatttcta	cagatatttc	acagatgaaa	1980
ggcaagcaat	tagagccttt	aggttggtca	aatcaaacia	agagctctgg	gccctgtgtc	2040
gggacgcctg	gtgttcctgg	ctggcctgca	ctgtcctgat	gcagcagatg	aagcgggaag	2100
aaaaactcac	actgacttcc	aagaccacca	caacctctcg	tctacattac	cttgcccagg	2160
ctctccaagc	tcagccattg	ggaccccagc	tcagagacct	ctgctctctg	gctgctgagg	2220
gcatctggca	aaaaaagacc	cttttcagtc	cagatgacct	caggaagcat	gggttagatg	2280
gggcccctct	ctccaccttc	ttgaagatgg	gtattcttca	agagcacccc	atccctctga	2340
gctacagctt	cattcacctc	tgtttccaag	agttctttgc	agcaatgtcc	tatgtcttgg	2400
aggatgagaa	ggggagagg	aaacattcta	attgcatcat	agatttggaa	aagacgctag	2460
aagcatatgg	aatacatggc	ctgtttgggg	catcaaccac	acgtttccta	ttgggacctg	2520
taagtgtatg	gggggagaga	gagatggaga	acatctttca	ctgccggctg	tctcagggga	2580
ggaaacctgat	gcagtgggtc	ccgtccctgc	agctgctgct	gcagccacac	tctctggagt	2640
ccctccactg	cttgtagtag	actcggaaca	aaacgttcct	gacacaagtg	atggcccatt	2700
tcaagaaaat	gggcagtgtg	gtagaacag	acatggagct	cttagtgtgc	actttctgca	2760
ttaaattcag	ccgccacgtg	aagaagcttc	agctgattga	gggcaggcag	cacagatcaa	2820
catggagccc	caccatggta	gtcctgttca	ggtgggtccc	agtcacagat	gcctattggc	2880
agattctctt	ctccgtcctc	aaggtcacca	gaaacctgaa	ggagctggac	ctaagtggaa	2940
actcgttag	ccactctgca	gtgaagagtc	tttgtaagac	cctgagacgc	cctcgtgccc	3000
tcttgagagc	cctgcggttg	gctggctgtg	gcctcacagc	tgaggactgc	aaggaccttg	3060
cctttgggct	gagagccaac	cagacctga	ccgagctgga	cctgagcttc	aatgtgctca	3120
cggatgctgg	agccaaacac	ctttgccaga	gactgagaca	gccgagctgc	aagctacagc	3180
gactgcagct	ggtcagctgt	ggcctcacgt	ctgactgtcg	ccaggacctg	gcctctgtgc	3240
ttagtgccag	ccccagcctg	aaggagctag	acctgcagca	gaacaacctg	gatgacgttg	3300
gcgtgcgact	gctctgtgag	gggtcaggc	atcctgectg	caaactcata	cgcttggggc	3360
tggaaccagac	aactctgagt	gatgagatga	ggcaggaact	gagggccctg	gagcaggaga	3420
aacctcagct	gctcatcttc	agcagacgga	aaccaagtgt	gatgacctct	actgagggcc	3480
tgtatgacgg	agagatgagt	aatagcacat	cctcactcaa	gcggcagaga	ctcgatcacg	3540
agagggcggc	ttcccatggt	gctcaggcta	atctcaaact	cctggacgtg	agcaagatct	3600
tcccaattgc	tgagattgca	gaggaaagct	ccccagaggt	agtaccgggt	gaactcttgt	3660
gcgtgccttc	tcttgcctct	caagggggacc	tgcatacgaa	gcctttgggg	actgacgatg	3720
acttctgggg	ccccacgggg	ctgtgggcta	ctgaggtagt	tgacaaagaa	aagaacttgt	3780
accgagttca	cttccctgta	gctggctcct	accgctggcc	caacacgggt	ctctgctttg	3840
tgatgagaga	agcggtgacc	gttgagattg	aattctgtgt	gtgggaccag	ttcctgggtg	3900
agatcaaccc	acagcacagc	tgatgggtgg	cagggcctct	gctggacatc	aaggctgagc	3960
ctggagctgt	ggaagctgtg	cacctccctc	actttgtggc	tctccaaggg	ggccatgttg	4020
acacatccct	gttccaaatg	gccactttta	aagaggaggg	gatgctcctg	gagaagccag	4080
ccaggggtgga	gctgcatcac	atagttcttg	aaaaccccag	cttctcccc	ttggggagtcc	4140
tcttgaataat	gatccataat	gccctgcgct	tcatccctgt	cacctctgtg	gtgttgcttt	4200
accacgcgct	ccatcctgag	gaagtcacct	tcacactcta	cctgatccca	agtgaactgt	4260
ccattcggaa	ggccatagat	gatctagaaa	tgaaattcca	gtttgtgcca	atccacaagc	4320
cacccccgct	gacccccact	tatatgggct	gtcgttacac	tgtgtctggg	tctgggttcc	4380
ggatgctgga	aatactcccc	aaggaaactg	agctctgcta	tcgaagccct	ggagaagacc	4440
agctgttctc	ggagtctctac	gttggccact	tgggatcagg	gatcaggctg	caagtgaag	4500
acaagaaaga	tgagactctg	gtgtgggagg	ccttgggtga	accaggagat	ctcatgcctg	4560
caactactct	gatccctcca	gcccgcatag	ccgtaccttc	acctctggat	gccccgcagt	4620
tgtgcacttt	tgtggaccag	tatcgagagc	agctgatagc	ccgagtgaca	tcgggtggag	4680
ttgtcttggga	caaactgcat	ggacaggtgc	tgagccagga	gcagtacgag	aggggtgctgg	4740
ctgagaacac	gaggcccagc	cagatgcgga	agctgttcag	cttgagccag	tcttgggacc	4800
ggaagtgcaa	agatggactc	taccaagccc	tgaaggagac	ccatcctcac	ctcattatgg	4860
aaactctggga	gaagggcagc	aaaaagggac	tcctgccact	cagcagctga	agtatcaaca	4920
ccagcccttg	accctttagt	cctggctttg	ctgacccttt	ctttgggtct	cagtttcttt	4980
ctctgcaaac	aagttgccat	ctggtttgcc	ttccagcact	aaagtaatgg	aacttatgat	5040
gatgccttgc	tgggcattat	gtgtccagcc	agggatgcac	agggggccca	gtcagggtgg	5100
ctacagcatc	tcagggatgt	cc				5122

<210> 30

<211> 1241

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6726656CB1

<400> 30
atztatggag gcaaccaaca aattacacag caatgaattc tggactgtag tcccacagcc 60
ttgagcctag atactagctg tgtcactaat actctgtgtg accaggaagc ttgagtggcc 120
aagggtttga taaatgtgga aaataaccgc tgttcgaata tctgttggc ctgctgtgat 180
gtgaggtgtg gaagaaacac gggagcagcc ttcctcagga cacctcttgt ttatctctct 240
agctctgaaa tcacatgaag ctgtggatgg agagtcacct gatagtccca gaaacccgtc 300
ccagcccaag gatgatgagt aaccagacgt tggtaaccga gttcatcctg cagggtttt 360
cggagcacc cagaataccg gtgttcttat tcagctgttt cctcttcctc tactctgggg 420
ccctcacagg taatgtcctc atcaccttgg ccatcacgtt caaccctggg ctccacgtc 480
ctatgtactt tttcttactc aacttggcta ctatggacat tatctgcacc tcttccatca 540
tgcccaaggc gctggccagt ctgtgtcgg aagagagctc catctcctac gggggctgca 600
tggcccagct ctatttctc acgtgggctg catcctcaga gctgctgctc ctcacgggtca 660
tggcctatga ccggtacgca gccatctgcc acccgctgca ttacagcagc atgatgagca 720
agggtgtctg cagcgggctg gccacagccg tgtggctgct ctgcccgtc aacacggcca 780
tcacacggg gctgatgctg cgcttgatt tctgtggccc caatgtcatt atccatttt 840
tctgcgaggt cctctccctg ctgcttctct cctgcagctc cacctacgtc aacgggtgtca 900
tgattgtcct ggcggtatgt ttctacggca tagtgaactt cctgatgacc atcgctcct 960
atggcttcat cgtctccagc atcctgaagg tgaagactgc ctggggagg cagaaagcct 1020
tctccacgtg ctcttccac ctcaccgtgg tgtgcattga ttacaccgt gtcttctacg 1080
cctacataag ccggtctct ggctacagcg caggaagag caagtggct ggctgtgt 1140
aactgtgtg gatcctacc ctcaacccc tcatctatac tttagaaac aaggaggtca 1200
aagcagcct caggaagctt ttccttttct tcagaaatta a 1241

<210> 31
<211> 1155
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472062CB1

<400> 31
atgaatgtcc ttctggcaga ttcaaattca aataaaaaga ttgtgcataa acacatctgc 60
agcctacagt cagcccccaa gactacgaac ctccaaccct caatatctga tatcctgcta 120
agtgttgaga gtaatgacag gaagaatgtg tctaagataa aaggggattg tttcaacaca 180
agagtatctt gtgattctaa aataacatcc atggagaata atacagagggt gattgaattc 240
atcctgcttg gtctaaccaa tgccccagaa ctacagggtc cctcttttat catgtttacc 300
ctcatctacc tcatcactct gactgggaac ctggggatga tcatattaat cctgctggac 360
tctcatctcc acactcccat gtacttttt ctcagtaacc tgtctcttgc aggcattggg 420
tactcctcag ctgtcactcc aaagggtttta actgggttgc ttatagaaga caaagccatc 480
tcctacagt cctgtgtgct tcagatgttc ttttgtgcag tctttgccac tgtggaaaat 540
tacctcttgt cctcaatggc ctatgaccgc tacgcagcag tgtgtaaccc cctacattat 600
accaccacca tgacaacacg tgtgtgtgct tgtctggcta taggctgtta tgtcattggg 660
tttctgaatg cttctatcca aattggagat acatttcgcc tctctttctg catgtccaat 720
gtgattcatc actttttctg tgacaaacca gcagtcatta ctctgacctg ctctgagaaa 780
cacattagtg agttgattct tgttcttata tcaagtttta atgtctttt tgcacttctt 840
gttaccttga tttcctatct gttcatattg atcaccattc ttaagaggca cacaggtaag 900
ggataaccaga agcctttatc tacctgtggg tctcacctca ttgccatttt cttattttat 960
ataactgtca tcattcatgta catacgacca agttccagtc attccatgga cacagacaaa 1020
attgcactctg ttctctacac tatgatcatc cccatgctca gtcctatagt ctataccctg 1080
aggaacaaag acgtgaagaa tgcattcatg aaggttgttg agaaggcaaa atattctcta 1140
gattcagttc ttttaa 1155

<210> 32
<211> 1260
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472067CB1

<400> 32

```
atgtggcag ctgcctttgc agactctaac tccagcagca tgaatgtgtc ctttgcacac 60
ctccactttg ccggagggtg cctgccctct gattcccagg actggagAAC catcatcccg 120
gctctcttgg tggctgtctg cctggtgggc ttctgtggaa acctgtgtgt gattggcacc 180
ctccttcaca atgcttggaaggaaagcca tccatgatcc actccctgat tctgaatctc 240
agcctggctg atctctccct cctgctgttt tctgcaccta tccgagctac ggcgtactcc 300
aaaagtgttt gggatctagg ctggtttgtc tgcaagtcct ctgactggtt tatccacaca 360
tgcatggcag ccaagagcct gacaatcggt gtggtggcca aagtatgctt catgtatgca 420
agtgacccag ccaagcaagt gagtatccac aactacacca tctggtcagt gctgggtggc 480
atctggactg tggctagcct gttaccctg ccggaatggg tcttttagcac catcaggcat 540
catgaagggtg tggaaatgtg cctcgtggat gtaccagctg tggctgaaga gtttatgtcg 600
atgtttggta agctctacc actcctggca ttgtgacctt cattattttt tgcagctttt 660
tatttctgga gagcttatga ccaatgtaaa aaacgaggaa ctaagactca aaatcttaga 720
aaccagatag tggctaaagca agtcacagt atgtgtctga gcattgccat catctctgct 780
ctcttgtggc tccccgaatg ggtagcttgg ctgtgggtat ggcattctga ggctgcaggc 840
ccggcccccac cacaaggttt catagccctg tctcaagtct tgatgttttc catctcttca 900
gcaaatcttc tcatttttct tgtgatgtcg gaagagttca gggaaggctt gaaagggtga 960
tggaaatgga tgataaccaa aaaacctcca actgtctcag agtctcagga aacaccagct 1020
ggcaactcag aggggtcttc tgacaagggt ccatctccag aatccccagc atccatacca 1080
gaaaaagaga aaccagctc tccctcctct ggcaaaggga aaactgagaa ggcagagatt 1140
cccatccttc ctgacgtaga gcagttttgg catgagaggg acacagtccc ttctgtacag 1200
gacaatgacc ctatccctg ggaacatgaa gatcaagaga caggggaagg tgttaaatag 1260
```

<210> 33

<211> 945

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472072CB1

<400> 33

```
atgggggatg tgaatcagtc ggtggcctca gacttcattc tgggtggcct cttcagtcac 60
tcaggatcac gccagctcct cttctccctg gtggctgtca tgtttgtcat aggccttctg 120
ggcaaacaccg ttcttctctt cttgatccgt gtggactccc ggctccatac acccatgtac 180
ttcctgctca gccagctctc cctgtttgac attggctgtc ccatggtcac catccccaag 240
atggcatcag actttctgcg gggagaagggt gccacctcct atggagggtg tgcagctcaa 300
atattcttcc tcacactgat ggggtgtggc gagggcgctc tgtttggtcct catgtcttat 360
gaccgttatg ttgctgtgtg ccagccctg cagtatcctg tacttatgag acgccaggta 420
tgtctgctga tgatgggctc ctctgggtg gtagggtgtc tcaacgcctc catccagacc 480
tccatcaccc tgcattttcc ctactgtgcc tcccgatttg tggatcactt cttctgtgag 540
gtgccagccc tactgaagct ctctgtgca gatacctgtg cctacgagat ggcgtgtc 600
acctcagggg tgctgatcct aatgtccct ctttccctca tcgccacctc ctacggccac 660
gtgttgccag ctgttctaag catgcgctca gaggaggcca gacacaaggc tgcaccacc 720
tgctcctcgc acatcacggt agtggggctc ttttatggtg ccgcccgtgt catgtacatg 780
gtccttgcg cctaccacag tccacagcag gataacgtgg tttccctctt ctatagcctt 840
gtcaccctta cactcaacc ccttatctac agtctgagga atccggagggt gtggatggct 900
ttggtcaaag tgcttagcag agctggactc aggcAAatgt gctga 945
```

<210> 34

<211> 765

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472074CB1

<400> 34

```
atggcctccc ggtatgtggc agtgggaatg atcttatcac agaccgtggg gggagtctctg 60
gggagcttct ctgttcttct ccattatctc tctttttact gcactgggtg cagggttaagg 120
tccacagatt tgattgttaa gcacctgatt gttagccaact tcttagctct ccgctgtaaa 180
ggagtcccc agacaatggc agcttttggg gttagatatt ttctcaatgc tcttgggtgc 240
```

```

aaacttgttt tctatctcca tagagtgggc aggggagtggt ccattggcac cacctgcctc 300
ttgagtgtct tccaggtgat cacggtcagc tccaggaaat ccaggtgggc aaaacttaaa 360
gagaaagccc ccaagcatgt tggcttttct gtctctctgt gctggatcgt gtgcatgttg 420
gtaaacatca tctttcccat gtatgtgact ggcaaatgga actacacaaa catcacagt 480
aacgaggatt tgggatactg ttctggggga ggcaacaaca aaatcgaca gacactgcgt 540
gcaatgttgt tatcattccc tgatgtgttg tgtctggggc tcatgtcttg ggtcagcagc 600
tccatggttt gcatcctgca caggcacaag cagcgggtcc agcacattga taggagcgat 660
ctctcccca gagcctcccc agagaacaga gctacgcaga gcatcctcat cctggtgagc 720
acctttgtgt cttcttacac tctctctgc cttttccaag tttga 765

```

<210> 35

<211> 1089

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472077CB1

<400> 35

```

atgtacaagg actgcatcga gtccactgga gactattttc ttctctgtga cgccgagggg 60
ccatggggca tcattctgga gtccctggcc atacttgga tctgtgtcac aattctgcta 120
ctcttagcat ttctcttct catgcgaaag atccaagact gcagccagt gaatgtcctc 180
cccaccagc tctcttctt cctgagtgtc ctggggctct tcggactcgc ttttgccttc 240
atcatcgagc tcaatcaaca aactgcccc gtacgtact ttctctttgg ggttctcttt 300
gctctctgtt tctcatgcct cttagctcat gcctccaatc tagtgaagct ggttcggggg 360
tgtgtctctt tctcctggac gacaattctg tgcattgcta ttggttgagc tctgttgcaa 420
atcattattt ccactgagta tgtgactctc atcatgacca gaggtatgat gttgtgaa 480
atgacaccct gccagctcaa tgtggacttt gttgtactcc tgggtctatgt cctcttctct 540
atggccctca cattcttctg ctccaaagcc accttctgtg gccctgtgtga gaactggaag 600
cagcatggaa ggctcatctt tatcactgtg ctcttctcca tcatcatctg ggtgggtgtg 660
atctccatgc tctgagagg caaccgcag ttccagcgac agccccagt ggacgacccg 720
gtcgtctgca ttgctctggt caccaacgca tgggttttcc tgctgctgta catcgtccct 780
gagctctgca ttctctacag atcgtgtaga caggagtgcc ctttacaagg caatgcctgc 840
ccgctcacag cctaccaaca cagcttccaa gtggagaacc aggagctctc cagagataaa 900
tggagggtct tactcaactc ggacttctta tcacacagt gtgcagccc agacagtgat 960
ggagctgagg aggatgtagc attaacttca tatggtactc ccattcagcc gcagactgtt 1020
gatccacac aagagtgttt catcccacag gctaaactaa gccccagca agatgcagga 1080
ggagtataa 1089

```

<210> 36

<211> 1334

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472082CB1

<400> 36

```

atgtgtaaat gcttcagaag tggcaatagc actccagtcc tgtgtcaccg aaactcagaa 60
gcatggcagc ccaggaaggc cccaagaaca cagcaaactg acatgggtta caccaattta 120
aattccaaga aagagtgcac gtacattaag gaaaatttca aaaagactgt tgacaagatc 180
gtggaccctg gaaaccattc ctcatgact gagtccattc tggctgggct ctcagaacag 240
ccagagctcc agctgcgcct ctctctctg ttcttaggaa tctgtgtggt cacagtgggt 300
ggcaacttgg gcatgatcac actgattggg ctcatgtctc acctgcacac acctatgtac 360
tatttctctc gcagtctgtc cttcattgac ttctgccatt ccactgtcat tacccttaag 420
atgctgggtga actttgcgac agagaagaac atcatctcct acctggaatg catggctcag 480
ctctatttat tcatgtttt tgctattgca gagtgtcaca tgttggctgc aatggcggtat 540
gactgttatg ttgccatctg cagccccttg ctgtacaatg tcatcatgtc ctatcaccac 600
tgcttctggc tcacagtggg agtttacatt ttaggcattc ttggatctac aattcatacc 660
agttttatgt tgagactctt tttgtgcaag actaatgtga ttaaccatta tttttgtgat 720
cttttccctc tcttggggct ctctgtctcc agcacctaca tcaatgaat actgggtctg 780
gtcttgagt catttaacat cctgatgcct gccttaacca tcttgcctt ttacatcttt 840
atcattgcca gcatcctccg cattcactcc actgagggca ggtccaaagc cttcagcact 900
tgagctccc acatcttggc tgttgcgtgt ttctttggat ctgcagcatt catgtacctg 960
cagccatcat ctgtcagctc catggaccag aggaaagtgt cgtctgtgtt ttatactact 1020

```

```

attgtgcccc tgctgaaccc cctgatctac agcctgagga ataaagatgt caaacttgcc 1080
gtgaagaaaa ttctgcatca gacagcatgt taatgaatag aatcaatgtt atgttggtac 1140
atcaagatag gtcttttggt tgattagata tctaacttat tggatttatt gttgagattt 1200
atgaaaattt agtgatgctc ttttatgtaa cacctctcca aaatattcct cgggtctgct 1260
tccatcgaac ttatatcca atgagcatat gtaaagaaat acaaagaata aaatcaaaag 1320
acttttgagg tttta
1334

```

<210> 37
 <211> 960
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472128CB1

```

<400> 37
atgacacctg gagaactagc ccttgccagt ggcaaccaca cccagtcac caagttcatc 60
ttgcagggat tctccaatta tccagacctc caggagcttc tcttcggagc catcctgctc 120
atctatgcca taacagtggg gggcaacttg ggaatgatgg cactcatctt cacagactcc 180
catctccaaa gccaatgta tttcttctc aatgtcctct cgtttcttga tatttggtac 240
tcttctgtgg tcacacctaa gctcttggtc aacttctctg tctctgacaa gtccatctct 300
tttgagggct gtgtggtcca gctcgcttc tttgtagtgc atgtgacagc tgagagcttc 360
ctgctggcct ccattggccta tgaccgcttc cttagccatct gtcaacccct ccattatggt 420
tctatcatga ccagggggac ctgtctccag ctggtagctg tgtcctatgc atttggtgga 480
gccaaactcg ctatccagac tggaaatgtc tttgccctgc ctttctgtgg gcccaaccag 540
ctaacaact actactgtga cataccacc cttctccacc tggcttgtgc caacacagcc 600
acagcaagag tggctcctcta tgtcttttct gctctggtca cccttctgcc tgcctgcagtc 660
attctcact cctactgctt ggtcttggtg gccattggga ggatgcgctc agtagcaggg 720
agggagaagg acctctccac ttgtgcctcc cactttctgg ccattgccat tttctatggc 780
actgtggttt tcacctatgt tcagcccat ggatctaata acaataccaa tggccaagta 840
gtgtccgtct tctacaccat cataattccc atgctcaatc ctttcatcta tagcctccgc 900
aacaaggagg tgaagggcgc tctgcagagg aagcttcagg tcaacatctt tcccggctga 960

```

<210> 38
 <211> 939
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472134CB1

```

<400> 38
atggacacag ggaactggag ccaggtagca gaattcatca tcttgggctt ccccatctc 60
cagggtgtcc agatttatct ctctctcttg ttgcttctca tttacctcat gactgtgttg 120
ggaaacctgc tgatattcct ggtggtctgc ctggactccc ggcttcacac acccatgtac 180
cactttgtca gcattctctc cttctcagag cttggctata cagctgccac catccctaag 240
atgctggcaa acttgetcag tgagaaaaag accatttcat tctctgggtg tctcctgcag 300
atctatttct ttcactccct tggagcgact gagtgctatc tcttgacagc tatggcctac 360
gataggatt tagccatctg ccggccctc cactacccaa ccctcatgac cccaacactt 420
tgtgcagaga ttgccattgg ctggttggtg ggaggcttgg ctgggccagt agttgaaatt 480
tccttgattt cagcctccc attctgtggc cccaatcgca ttcagcacgt ctttctgtac 540
ttccctctctg tgetgagttt ggcttgcaat gatacgtcta taaatgtcct agtagatttt 600
gttataaatt cctgcaagat cctagccacc ttctgtctga tctctgctc ctatgtgcag 660
atcatctgca cagtgtcag aattccctca gctgccggca agaggaaggc catctccacg 720
tgtgcctccc acttcaactgt ggttctcact ttctatggga gcaccccttc catgtatgtg 780
cagctgaaga agagctactc actggactat gaccaggccc tggcagtggt ctactcagtg 840
ctcacaccct tctcaaccc cttcatctac agcttgcgca acaaggagat caaggaggct 900
gtgaggaggc agctaaagag aattgggata ttggcatga
939

```

<210> 39
 <211> 968
 <212> DNA
 <213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472136CB1

<400> 39

```

ggatgaacca gactttgaat agcagtggga cagtggagtc agccctaaac tattccagag 60
ggagcacagt gcacacggcc tacctgggtc tgagctccct ggccatgttc acctgcctgt 120
gcgggatggc aggcaacagc atggtgatct ggctgctggg ctttcgaatg cacaggaacc 180
ccttctgcat ctatatcctc aacctggcgg cagccgacct cctcttctc ttcagcatgg 240
cttccacgct cagcctggaa acccagcccc tggtaatac cactgacaag gtccacgagc 300
tgatgaagag actgatgtac tttgcctaca cagtgggcct gagcctgtg acggccatca 360
gcacccagcg ctgtctctct gtccctcttc ctatctggtt caagtgtcac cggcccaggc 420
acctgtcagc ctgggtgtgt ggctgtgtgt ggacactctg tctcctgatg aacgggttga 480
cctcttctct ctgcagcaag ttcttgaaat tcaatgaaga tcggtgcttc aggggtggaca 540
tgatcaagcg cgccctcatc atgggggtct taacccagtg gatgactctg tccagcctga 600
cctcttttgt ctgggtgcgg aggagctccc agcagtggcg gcggcagccc acacggctgt 660
tcgtggtggt cctggcctct gtccctggtgt tccctcatctg tccctgcct ctgagcatct 720
actggtttgt gctctactgg ttgagcctgc cgcccgagat gcaggtcctg tgcttcagct 780
tgatcagcct ctccctctcc gtaagcagca gcgccaacc cgctcatctac ttcctgggtg 840
gcagccggag gagccacagg ctgcccacca ggtccctggg gactgtgctc caacaggcgc 900
ttcgcgagga gcccagagctg gaaggtgggg agacgccac cgtgggcacc aatgagatgg 960
gggcttga

```

<210> 40

<211> 1000

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472142CB1

<400> 40

```

attaaatgaa gggaatagta ctggaagaaa tatagatgaa agaaagaaaa tgcaaggaga 60
aaacttcacc atttggagca ttttttctt ggagggattt tcccagtacc caggggttaga 120
agtggttctc ttctgttca gccttgtaat gtatctgaca acgctcttgg gcaacagcac 180
tcttattttg atcactatcc tagattcacg ccttaaaacc cccatgtact tattccttgg 240
aaatctctct ttcatggata tttgttacac atctgcctct gttcctactt tgctggtgaa 300
cttgcgtgtca tcccagaaaa ccattatctt ttctgggtgt gctgtacaga tgtatctgtc 360
ccttgccatg ggctccacag agtgtgtgct cctggccgtg atggcatatg accgttatgt 420
ggccatttgt aacccgctga gatactccat catcatgaac aggtgctgtc gtgcacggat 480
ggccacggct tcctgggtga cgggttgccct gaccgctctg ctggaaacca gttttgccct 540
gcagataccc ctctgtggga atctcatcga tcacttcacg tgtgaaatc tggcgggtgct 600
aaagttagct tgcacaagtt cactgctcat gaacaccatc atgctggtgg tcagcattct 660
cctcttgcca attccaatgc tcttagtttg catctcttac atcttcatcc ttccactat 720
tctgagaatc acctcagcag agggaagaaa caaggctttt tctacctgtg gtgccattt 780
gactgtggtg attttgtatt atggggctgc cctctctatg tacctaagc cttcttcatc 840
aaatgcacaa aaaatagaca aaatcatctc gttgctttac ggagtgctta cccctatggt 900
gaaccccata atttacagtt taagaaacaa ggaagtcaaa gatgctatga agaaattgct 960
gggcaaaata acattgcatc aaacacacga acatctctga

```

<210> 41

<211> 1008

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472171CB1

<400> 41

```

ctgtggacca tctcttcaga actctgcagc atggagccgc tcaacagaac agaggtgtcc 60
gagttctttc tgaaaggatt ttctggtcac ccagccctgg agcatctgct cttccctctg 120
tgctcagcca tgtacctggg gaccctcctg gggaaacacag ccatcatggc ggtgagcgtg 180
ctagatatcc acctgcacac gcccggtgac ttcttctctg gcaacctctc taccctggac 240
atctgtcaca cgccacctt tgtgcctctg atgctggtcc acctcctgtc atcccgaag 300
acctctcct ttgctgtctg tgccatccag atgtgtctga gcctgtccac gggctccacg 360
gagtgcctgc tactggccat cacggcctat gaccgctacc tggccatctg ccagccactc 420

```

```

aggtagccag  tgctcatgag  ccaccggctc  tgcgtgctgc  tgatgggagc  tgcctgggctc  480
ctctgcctcc  tcaagtcggg  gactgagatg  gtcattctcca  tgaggctgcc  cttctgtggc  540
caccacgtgg  tcagtcactt  cactgcaag  atcctggcag  tgctgaagct  ggcatgcggc  600
aacacgtcgg  tcagcgaaga  ctctctgctg  gcgggctcca  tcctgctgct  gcctgtacct  660
ctggcattca  tctgcctgtc  ctacttgctc  atcctggcca  ccactctgag  ggtgccctcg  720
gccgccaggt  gctgcaaagc  cttctccacc  tgcttggcac  acctggctgt  agtgctgctt  780
ttctacggca  ccattcatctt  catgtacttg  aagcccaaga  gtaagggaagc  ccacatctct  840
gatgaggtct  tcacagtcct  ctatgccatg  gtcacgacca  tgctgaacct  caccatctac  900
agcctgagga  acaaggaggt  gaaggaggcc  gccaggaagg  tgtggggcag  gagtccggcc  960
tccagtgagg  gagggcgggg  ctctgtacag  acgcaggctc  cagggttag  1008

```

<210> 42

<211> 972

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472172CB1

<400> 42

```

ggaaaccctg  ccaccatat  agtagttgtc  atgggaaact  ggagcactgt  gactgaaatc  60
accctaattg  ccttcccagc  tctctgggag  attcgaatat  ctctcttcgt  ggttcttgtg  120
gtaacttaca  cattaacagc  aacaggaaac  atcaccatca  tctccctgat  atggattgat  180
catgcctgct  aaactccaat  gtacttcttc  ctacagtaatt  tgctctttct  ggatatctta  240
tacaccactg  tcattacccc  aaagtgtgtg  gcctgcctcc  taggagaaga  gaaaaccata  300
tcttttgctg  gttgcatgat  ccaaacatat  ttctacttct  ttctggggac  ggtggagttt  360
atcctcttgg  cgggtgatgt  ctttgaccgc  tacatggcta  tctgcgacct  actgcactac  420
acggtcatca  tgaacagcag  ggcctgcctt  ctgctgggtc  tgggatgctg  ggtgggagcc  480
ttctgtgtct  tggtgtttcc  aaccattgta  gtgacaaggc  taccttactg  taggaaagaa  540
attaatcatt  tcttctgtga  cattgcccct  cttcttcagg  tggcctgtat  aaatactcac  600
ctcattgaga  agataaaact  tctcctctct  gcccttgta  tcttgagctc  cctggcattc  660
actactgggt  cctacgtgta  cataatttct  accatcctgc  gtatcccctc  caccagggc  720
cgtcagaaaag  ctttttctac  ctgtgtctct  cacatcactg  ttgtctccat  tgcccacggg  780
agcaacatct  ttgtgtatgt  gagaccaat  cagaactcct  cactggatta  tgacaagggt  840
gccgctgtcc  tcattcacagt  ggtgacctct  ctctgaacc  cttttatcta  cagcttgagg  900
aatgagaagg  tacaggaagt  gttgagagag  acagtgaaca  gaatcatgac  cttgatacaa  960
aggaaaactt  ga  972

```